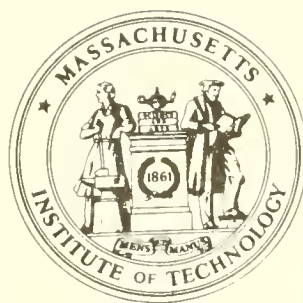




LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

COPY 1

EMPIRICAL FOUNDATIONS

OF

ECONOMIC ANALYSIS

83-64

Geoffrey P.E. Clarkson

MASS. INST. TECH.

OCT 30 1964

DEWEY LIBRARY

This essay is a draft for private circulation and comment. It should not be cited, quoted, or reproduced prior to publication without the written consent of the author. The research is supported by a grant from the Sloan School of Management from the funds provided by the Ford Foundation and NASA for research in business finance and organizational decision behavior respectively.

HD28
.M414
no. 83-64

Dewey

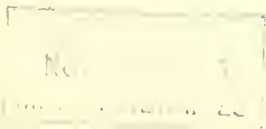
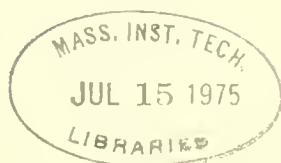


TABLE OF CONTENTS

	PAGE
 <u>PART ONE: ECONOMICS AND THE STRUCTURE OF A SCIENTIFIC SYSTEM</u>	
1. INTRODUCTION	1
2. ON THE STRUCTURES AND CHARACTERISTICS OF A SCIENTIFIC THEORY	12
1. The Structure of a Theory, 13 2. Interpreted Systems and Empirical Science, 19 3. Testing the Interpreted Systems, 22 4. Scientific Explanation and Prediction, 26	
 <u>PART TWO: MATHEMATICAL FOUNDATIONS OF CLASSICAL ANALYSIS</u>	
3. CLASSICAL FOUNDATIONS OF ECONOMIC ANALYSIS	31
1. The Technique of Equilibrium Solutions, 37 2. Comparative Statics and the Displacement of Equilibrium, 40	
4. THE THEORY OF MARKET EQUILIBRIUM	50
1. Conditions and Postulates of the Theory, 51 2. External Economies and the Slope of the Supply Function, 57 3. Market Equilibrium, 63 4. Static and Dynamic Stability of Equilibrium, 70 5. General Market Equilibrium, 79	
5. THE EMPIRICAL CONTENT OF MARKET THEORIES	85
1. Concepts and Conditions, 85 2. Testing and Theory's Hypotheses, 95	
6. AN EMPIRICAL ANALYSIS OF THE CLASSICAL DEDUCTIVE SYSTEM . .	107
1. The Basic Deductive System, 108 2. The General Deductive System, 112 3. The Market Conditions, 118 4. Equilibrium Analysis and Economic Theory, 120	
 <u>PART THREE: FOUNDATIONS AND CHARACTERISTICS OF ECONOMETRIC ANALYSIS</u>	
7. FOUNDATIONS OF ECONOMETRIC ANALYSIS	128
1. The Basic Postulates, 130 2. Estimating the Parameters, 136 3. Testing Statistical Hypotheses, 149	

8.	SOME PROBLEMS OF APPLICATION	158
	1. Lagged Endogenous Variables, 158 2. Simultaneous Interdependence, 160 3. The Identification Problem, 164 4. Forecasting, 175	
9.	THE EMPIRICAL CONTENT OF ECONOMETRIC THEORY	180
	1. The Initial Conditions, 180 2. Testing Econometric Hypotheses, 192	
10.	EXPLANATION AND PREDICTION IN ECONOMETRICS	200
	1. Prediction and Empirical Tests, 202 2. Econometric Forecasts as Predictive Tests, 204 3. Ceteris Paribus and Population Stability, 208 4. Micro-Analysis and Statistical Tests, 212	

PART FOUR: BEHAVIORAL FOUNDATIONS OF ECONOMIC ANALYSIS

11.	THE ECONOMIST'S PERSPECTIVE	221
	1. Economic Analysis and the Problem of a River Process, 225 2. Classical Analysis and Decision Processes, 238	
12.	FOUNDATIONS OF BEHAVIORAL THEORY	242
	1. A Theory of Individual Decision Behavior, 244 2. Goals and the Structure of Decision Processes, 248 3. On the Explanation of Decision Behavior, 255 4. Towards a General Theory of Decision Behavior, 260	
13.	SOME PROBLEMS OF APPLICATIONS	264
	1. On the Construction of Decision Theories, 265 2. On Testing Decision Theories, 272 3. The Problem of Errors, 277 4. Heuristics, Algorithms and Statistical Tests, 285	
14.	THEORIES OF ECONOMIC BEHAVIOR--THE CONSUMER AND THE FIRM . . .	291
	1. On a Theory of Consumer Behavior, 295 2. On a Theory of the Firm, 307	

	PAGE
15. TOWARDS A THEORY OF MARKET BEHAVIOR	318
1. Price Behavior in a Security Market, 321	
2. Testing the Market Processes, 335	
16. BEHAVIORAL THEORY, MICRO-ANALYSIS, AND POLICY DECISIONS . .	339
1. Behavioral Theories and Econometrics, 342	
2. Time Series and Behavioral Analysis, 350	
3. Behavioral Theory and Policy Decisions, 354	
17. TOWARDS A SCIENCE OF ECONOMICS	359

PART ONE

Economics and the Structure

of a

Scientific System

Chapter 1

INTRODUCTION

Economics is not a well specified discipline. At a general level, it can be defined as the social science which takes as its subject matter the behavior of individuals or groups of individuals engaged in the production, exchange, and consumption of goods and services. The investigation of such behavior has led some economists to explore the boundaries between economics and psychology, while others have been concerned with the frontiers conjoining economics and sociology as well as other social sciences. In brief, the substance of economics is expanding and the expediency of these alterations is often a subject of debate. Throughout such vicissitudes the central objective of economics remains unchanged. It is to describe and explain the behavior of several types of homoeconomics.

This book is concerned with the science of economics. It is not a treatise on the methods of science. Rather it is an inquiry into the conditions under which a science of economics can be developed. While much of the economic literature is devoted to the many aspects of public and private policy decisions, such prescriptions must ultimately be based upon a knowledge of the relevant economic behavior. The acquisition of knowledge is the task of science. And it is toward an examination of this process in economics that this book is directed.

The recognition of a specific problem always indicates the beginning of a scientific venture. Success is achieved when an answer is attained. A scientific investigation is distinguished by the form of the answer as well as the type of question to which the venture is addressed. Problems are raised by posing the question "Why?" and are resolved if the answer is in the form of an explanation of the "Why?". To only answer the "What?", "When?", or "How?" is to evade the primary task of science. The second stage of the process requires constructing a theory about the events or difficulty in question. The theory must be stated in such a way that it can be employed as part of the answer to the "Why?". If the theory is successful in this regard, i.e., is able to provide the necessary explanatory link, then this theory is accepted as part of our empirical knowledge about the phenomena in question.

For example, on a cold, rainy day we frequently observe moisture forming on the inside of automobile windows. The question might well be put, "Why does moisture form on the inside of automobile windows under these conditions?" The answer entails employing our knowledge of the behavior of water vapor in air to explain the observed moisture. The explanation itself, in outline form, would consist of the following types of statements: First of all there would be observations recording that the temperature of the glass in contact with the cold, outside air is considerably lower than the temperature of the air within the automobile, and that the air within the automobile contains water

vapor. Second, there would be the relevant empirical hypotheses which assert that water vapor in air precipitates as a liquid whenever the air comes in contact with a surface that is sufficiently cold.

From these statements an explanation of the "Why?" of the moisture on the window can be constructed, and the general statements or theory within this explanation are regarded as part of our knowledge about the behavior of water in air.

In order to provide answers to specific problems the scientific process leads us to construct theories about particular phenomena. The theories, in turn, if they survive empirical tests, are then the basis of what constitutes our empirical knowledge of those particular kinds or classes of events. Consequently, the scientific venture, while principally directed toward providing answers to specific problems, is at the same time the process by which we acquire empirical knowledge about specific phenomena.

As economists, therefore, we are faced by a two-fold task: to explain the occurrence of specific economic phenomena; and at the same time to acquire and develop a body of empirical knowledge about these economic events. As long as our theories are responsive to data and can be confuted by experimental test, then those theories that survive will allow us to perform both of these tasks.

The heart of the matter is whether the theories of economics are responsive to data and whether they can be submitted to a process of refutation by empirical test. For if it should be the case that economic theories are stated in such a way as to preclude their disconfirmation by empirical test, then it follows that we do not have a basis from which to develop a body of empirical knowledge of economic events. While such a conclusion may not perturb some economists, particularly those who are persuaded that a science of human activity cannot be developed,^{1/} it is profoundly disturbing to those who accept the notion that the principal task of economists is to acquire scientific knowledge of economic behavior.

Of course, there are many economists who are primarily interested in deriving normative solutions to the multitude of economic issues which face nations, communities, firms, and individuals. Such solutions are stated in terms of what ought to be done under specific circumstances to achieve certain objectives. These prescriptions are not, as a rule, produced from out of thin air. They are deduced as consequences of particular theoretical frameworks. If the theory itself is a testable statement of certain economic processes then such policy conclusions as are derived from it are founded

^{1/} A lucid statement of this position is to be found in: Ludwig Von Mises, Human Action: A Treatise on Economics, Yale University Press, 1949.

The heart of the matter is whether the theories of economics are responsive to data and whether they can be submitted to a process of refutation by empirical test. For if it should be the case that economic theories are stated in such a way as to preclude their disconfirmation by empirical test, then it follows that we do not have a basis from which to develop a body of empirical knowledge of economic events. While such a conclusion may not perturb some economists, particularly those who are persuaded that a science of human activity cannot be developed,^{1/} it is profoundly disturbing to those who accept the notion that the principal task of economists is to acquire scientific knowledge of economic behavior.

Of course, there are many economists who are primarily interested in deriving normative solutions to the multitude of economic issues which face nations, communities, firms, and individuals. Such solutions are stated in terms of what ought to be done under specific circumstances to achieve certain objectives. These prescriptions are not, as a rule, produced from out of thin air. They are deduced as consequences of particular theoretical frameworks. If the theory itself is a testable statement of certain economic processes then such policy conclusions as are derived from it are founded

^{1/} A lucid statement of this position is to be found in: Ludwig Von Mises, Human Action: A Treatise on Economics, Yale University Press, 1949.

upon scientific grounds. But if the theory cannot be confuted by empirical test, there is no possible way by which the stated prescriptions can be empirically shown to lead to the desired results.

Consider, for example, the case of a gunner in charge of a battery of field artillery who desires to shell a specific target. From a theory of the behavior of projectiles and from a knowledge of the muzzle velocity of the shells it is possible to compute, under observed wind conditions, the requisite elevation and bearing. Accordingly, to strike the target the gunner can be instructed to elevate his guns to such-and-such a position and align them onto a bearing of so many degrees.

Under certain conditions it is not necessary to start with any theory at all. If the target is both visible and stationary the gunner can be instructed to arrive at the correct position for his guns by trial and error. However, if the target is neither stationary nor directly visible, the only sound way to proceed is by employing the theory of projectiles in conjunction with whatever information is available about the target's behavior.

The situation in economics is not entirely dissimilar. If the world remained the same while experiments were conducted with various economic policies, it would not be long before a set of normative statements were developed which could be applied to a large range of problems under specified conditions. Unfortunately, the economist's world is not nearly so convenient as the gunner's; it does not remain the same while various policies are tried out. Moreover, in most situations of interest it is seldom clear, even after

the event, why in fact the observed behavior took place as it did. If economists are to be able to produce normative statements or policy prescriptions on request, these assertions must be directly derivable from the relevant body of theory. And, what is even more important, if we are to be capable of demonstrating why the adoption of a specific policy will lead to the occurrence of a particular outcome or end result, then the theory upon which the policy is based must be confutable by empirical test. Therefore, whether one adopts the perspective of the scientist, the advisor, or the policy maker it is clear that the economist's primary objective is to acquire a body of empirical knowledge about economic phenomena.

In a recent investigation of the theory of consumer demand^{2/} it is shown that this particular theory cannot be submitted to a process of refutation by empirical test. That is to say, it is demonstrated that while data can always be found to support many of the statements in the theory, the theory, including all propositions derivable from it, does not have sufficient empirical content to allow it to be confuted by experimental test. Since a theory must be refutable if it is to serve as part of the explanatory process, it follows directly that we are unable to employ the theory of consumer demand to explain consumer behavior. To acquire empirical knowledge about particular events we

^{2/} G. P. E. Clarkson, The Theory of Consumer Demand: A Critical Appraisal, Prentice-Hall, Englewood Cliffs, 1963.

must have theories that can withstand empirical tests. Consequently, an untestable theory of demand does not provide us with a theory with which we can acquire empirical knowledge about the behavior of consumers.

As scientists and economists we ought to be alarmed by this state of affairs. If one specific theory places us in a position where we are unable either to explain the occurrence of economic events or to acquire empirical knowledge about a particular set of economic phenomena, where are we with respect to the remainder of microeconomic theory? Is one theory beyond the pale of empirical science while the rest are able to satisfy scientific criteria? Or is it the case that the theories of microeconomics are sufficiently dependent on one another so that if one cannot be confuted then this dearth of empirical content is passed on to the remaining body of theory? Alternatively, since all microeconomic theory can be generated from certain mathematical foundations with the addition of a few behavioral assumptions, is it the method by which these theories are developed and stated which leads to their untestable condition?

These are not idle questions. Indeed, it is part of the purpose of this essay to inquire into the nature of their answers. In particular the second part of the book is devoted to an examination of the last of these questions. For if it can be shown that the method by which microeconomic theory is developed -- i.e. its mathematical foundations -- leads inevitably to the construction of untestable economic theories then such a conclusion will simultaneously provide answers to the first three questions.

However, in order to be able to discover the answers to these questions one needs to identify the criteria by which an analysis can be conducted into the empirical properities of microeconomic theories. Further, it is necessary to know the criteria by which a theory can be classified as being able to be refuted by empirical test. While I have no desire to compete with philosophers of science, the reader and the author must agree, for the duration of the book at least, on what is meant by each particular term. As a result, it is necessary to examine, however briefly, the nature of scientific theory as well as the criteria these theories are expected to meet.

With a description of the analytical tools -- the object of the next chapter -- one can then proceed directly to an examination of the mathematical foundations of microeconomics. If the foundations permit the development of theories which satisfy the criteria of empirical science, then one can conclude that the non-testability of the theory of consumer demand is likely to be an isolated case in microeconomic theory. If, on the other hand, the mathematical foundations are such that they inevitably lead to the development of economic theories which cannot be confuted by empirical test, then the theory of consumer behavior is no longer an isolated phenomenon. Indeed, in this case it is the foundations themselves which are the primary source of the empirical difficulties.

The second part of this book, therefore, is devoted to the task of examining the mathematical foundations of microeconomics. Since these foundations do not by themselves constitute a theory about observable phenomena, but are rather

the basis from which microeconomic theories are developed, the analysis in these chapters includes some specific examples of these micro theories. Accordingly, the object of this inquiry is to examine both the empirical content of these theories and of the foundations from which they are derived.

For one moment consider one of the possible outcomes of such an investigation. In particular, what if the analysis demonstrates that microeconomic theories, by virtue of the manner in which they are derived and expressed, cannot be subjected to a process of disconfirmation by empirical test? In this situation it might well be argued that it surely is possible to treat some microeconomic propositions as independent statements and subject them to test by themselves. In other words, if microeconomic theories cannot themselves be subjected to test, can one take individual hypotheses and by a specific process of parameter estimation independently subject them to empirical tests?

One answer to this query is given by the procedures for measurement and parameter estimation which are a part of the process called econometric analysis. If by the application of econometrics either single hypotheses or groups of such hypotheses can be confuted, then it will no longer matter whether the classic, mathematically derived, microeconomic theories can be disconfirmed. For, once one or more hypotheses can be independently subjected to test they can stand upon their own empirical feet and, as a consequence, are no longer dependent upon their ancestry.

The question posed by econometrics, which is the subject of the third part of this essay, is as follows: are econometric methods sufficient to permit the disconfirmation of economic hypotheses by empirical test? If the answer is in the affirmative the search for the foundations of a science of economics is at an end. Once hypotheses exist that can survive a direct confrontation with empirical tests these very propositions become the corner stone upon which testable theories can be built.

But if, as the analysis in the third segment of the book shows, current econometrics methods are not sufficient to permit the disconfirmation of econometric hypotheses then it is necessary to search among more recent proposals in behavioral theory for the possible existence of the required empirical foundations. Accordingly, the final part of the book is devoted to an analysis of recent attempts to reconstruct economic theory on behavioral grounds.

The main task facing economists is to acquire empirical knowledge of economic phenomena. In order to succeed in this endeavor it must be possible to develop testable theories of economic processes. Although this entire book is devoted to a search for the requisite empirical foundations, it is not until the final chapters that specific grounds are described upon which a science of economics can be built. This, of course, is not to say that these foundations are the only possible ones capable of performing in the required manner. Clearly, after one solution to a problem is presented other more efficient or elegant ones may rapidly follow. The point to note, however, is that by themselves

neither the classic, mathematical formulations nor those derived from econometrics can stand by themselves as bases from which refutable propositions can be derived.

To demonstrate the validity of this last assertion it is necessary, as has already been noted, to take a brief foray into the nature and characteristics of scientific theories. While such a discussion may appear superfluous to the reader, in my opinion it is essential to the progress and force of this indagation to clarify beforehand the terms, concepts and analytical framework that is to be used.

Chapter 2

ON THE STRUCTURE AND CHARACTERISTICS OF SCIENTIFIC THEORY

Everyone who has ever tried to devise an answer to the question, "Why does (did) such-and-such an event occur?" has been in the position of trying to construct a sequence of steps which leads in a logical fashion from some starting point to the event in question. Frequently, if the event is of sufficient interest some of the individual steps in this process will be stated as specific hypotheses or propositions about the behavior of the events under consideration. These hypotheses will refer to certain properties or characteristics of these events. And when such a collection of propositions forms a coherent whole they are referred to collectively as a theory.

Not all such theories, whatever language they are stated in, are adjudged to be a part of science. Indeed, the problem of what distinguishes scientific theories from other forms of theory has been an issue of central concern to philosophers of science for a large number of years.^{1/} Accordingly it is not altogether surprising to discover that there are a number of issues upon which philosophers continue to disagree. As a result, it is clearly not possible to

^{1/} An excellent and comprehensive statement of the views of a specific group of philosophers of science is to be found in: O. Neurath, R. Carnap, and C. Morris, (eds.), International Encyclopedia of Unified Science, University of Chicago Press, Vols. I and II, 1955.

assert that there is only one acceptable point of view or one set of criteria. But, on the other hand, in order to perform a theoretical analysis such as is proposed in this book a single, analytical framework is required. To satisfy this acquirement as well as to circumvent a number of the thorny, undecided issues, I have resolved the problem by accepting and now presenting what appears to me to represent a consensus of philosophic opinion.

1. The Structure of a Theory^{2/}

A theory may be considered as a set of propositions which are expressed in terms of a particular vocabulary. The vocabulary of the theory in turn consists of two sets of terms. The first set contains all the logical terms, e.g. is, not, or, implies, if and only if, etc. which the theory contains. The other set consists of the extra logical terms, some of which may or may not refer directly to observables. In either case the extra logical terms of the vocabulary can usually be divided into two classes: those for which no definition is specified within the theory--the so-called primitive terms; and those which are defined in terms of the primitive and/or other concepts in the vocabulary. In a similar manner the propositions of a theory can also be divided into two classes: The first class contains those sentences

^{2/}A lucid exposition of the structure and characteristics of scientific theory is to be found in: R.B. Braithwaite, Scientific Explanation, Cambridge University Press, London, 1953, Chapter 2.

which are primitive or basic to the development of the theory. These sentences are frequently called the postulates or axioms of the theory. The second class consists of those sentences which are derived by the rules of logical inference from the conjunction of the defined terms and the postulates. While this manner of representing the structure of a theory is in some what formal terms it provides a convenient schema by which they may be described. Hence, when referring to a theory we can conceive of it as consisting first of all of its primitive terms and its primitive propositions or postulates. From the primitive terms the remainder of the vocabulary can be developed by the definitional rules contained in the theory. From the concatenation of the complete vocabulary and the postulates the entire theory can be developed by the selective application of the canons of deductive logic.

Viewed in this manner, a theory is a deductive system. Once the primitive terms and postulates are specified the derivation of further propositions or theorems can be carried out solely by the application of the relevant rules of inference. At no time during this deductive process is it necessary to refer to the meaning of the sentences being derived. As long as their derivation follows the requisite logical rules within a specified deductive framework their validity as sentences within the theory is not dependent on any meaning which might be ascribed to them.

For example, consider a simple deductive system^{3/} which contains one

^{3/}I am indebted for this example to H. Hochberg, "Axiomatic Systems, Formalization, and Scientific Theories," in L. Gross, Symposium on Sociological Theory, Row, Peterson and Co., pp. 407-436.

primitive term and four primitive sentences or postulates;

The primitive term in this system is 'point'. The postulates are:

P-1: Every line has at least two points as members

P-2: There are at least two lines

P-3: Two lines do not have two points in common

P-4: For any two points there is a line that they are both members of.

The one defined term in these postulates is 'line', and it is defined in terms of 'point' as follows:

D-1: 'line' for 'class of points'

Given the vocabulary and the postulates we can now derive further propositions within this deductive system as follows:

Step 1: There are two lines--call them A and B By P-2

Step 2: A has two points, a₁ and a₂; By P-1

B has two points, b₁ and b₂. and Step 1

Step 3: At most one point of A can be identical with By P-3 and
one point of B. Hence there are three points. Step 2.

As a result of this sequence of inferences we can now add the following proposition or theorem to our system:

T-1: There are at least three points.

Clearly, we can proceed to add further propositions in a similar manner. We can begin with T-1 and by the application of P-4 and the definition of a 'line' arrive at the proposition T-2: There are at least three lines. Throughout this process no reference has been made to the empirical meaning of T-1 and T-2 or P-1, P-2, P-3, P-4 and such reference is quite unnecessary when we are examining the theory as a deductive system.

For example, we could have replaced 'point' and 'line' by the Greek letters ' Σ ' and ' Λ ' so that the definition and postulates would now read:

D'-1: ' Λ ' for 'class of Σ '

P'-1: Every Λ has at least two Σ as members.

P'-2: There are at least two Λ .

P'-3: Two Λ do not have two Σ in common.

P'-4: For any two Σ there is a Λ that they are both members of.

Clearly, substituting Greek letters for the terms 'point' and 'line' is not going to affect our ability to deduce T'-1: There are at least three Σ , or T'-2: There are at least three Λ . Nor would the deductions be affected if the logical words in the postulates were replaced by other formal symbols. In fact, if this were done the result would be a deductive system of the type logicians prefer to deal with. A system of this sort is called an uninterpreted calculus or deductive system.^{4/}

The point to note is that the process of deductive inference is governed by a specific set of rules which makes no reference to what the propositions themselves refer to. In order to determine what is meant by a specific postulate or proposition within a deductive system an additional procedure is employed.

To explicate this procedure suppose that L, M, N, Q, are the names of propositions in a calculus where this calculus consists of simple English sentences. Then, if P, Q, R, S are variables in the language used to talk about the calculus (the metalanguage), these variables can

^{4/} The properties of a number of such calculi are examined in detail in: A. Church, Introduction to Mathematical Logic, Princeton University Press, Vol. 1, 1956.

denote any of the sentences in the deductive system. For example, if the calculus consists of sentences like 'this object is made of iron' and 'this object expands when heated' their names are L and M respectively; but in the metalanguage they are referred to as the variables P and Q.

To complete the procedure for determining what is meant by such statements in a calculus, an interpretation must be given to the logical terms that are employed to link individual sentences together to form new propositions. For example, if a logical connective is represented in the metalanguage by ' $\rightarrow$ ' then the problem to be resolved is to decide on a procedure for determining what is being asserted by the statement 'this object is made of iron' ' $\rightarrow$ ' 'this object expands when heated'.

The resolution proceeds in the following way. Each statement within the calculus can be assigned a truth value--i.e. P must either have the value 'true' or 'false'. At the same time each logical connective must have an associated truth table, where the truth table specifies the truth value of a statement given the different possible combinations of the truth values of its component parts. To illustrate this point the truth tables for two connectives ' $\rightarrow$ ' and ' $\sim$ ' are shown in Tables (I) and (II).

Table (I)

<u>P</u>	<u>Q</u>	<u>P $\rightarrow$ Q</u>
<u>T</u>	<u>T</u>	<u>T</u>
<u>T</u>	<u>F</u>	<u>F</u>
<u>F</u>	<u>T</u>	<u>T</u>
<u>F</u>	<u>F</u>	<u>T</u>

Table (II)

<u>P</u>	<u>$\sim$P</u>
<u>T</u>	<u>F</u>
<u>F</u>	<u>T</u>

From Table (I) it is clear that the meaning of the symbol ' $\rightarrow$ ' corresponds to the meaning usually accorded to the English words 'if...then' or 'implies', and from Table (II) the meaning of the symbol ' $\sim$ ' corresponds to the accepted usage of the word 'not'. But the point to note is that no matter what the truth values are of the component parts, the truth table for the relevant connective defines the truth value for the entire statement.

Consider, for example, the statement $\underline{P} \rightarrow \underline{Q}$ where the truth value of $\underline{P}$ is 'true' and the value of $\underline{Q}$ is 'false'. Manifestly the value of $\underline{P} \rightarrow \underline{Q}$ is 'false'. Further, this is the only ordering of the truth values of $\underline{P}$ and $\underline{Q}$ which will result in the statement $\underline{P} \rightarrow \underline{Q}$ having the value 'false.'

Within any calculus there are some statements which have the value 'true' for all truth values of their component parts. Such propositions are called tautologies and are usually described as being logically true. If all propositions within a deductive system are tautologies then this is called a 'pure'^{5/} or uninterpreted deductive system. On the other hand, if the truth values of some sentences depend not only upon the values of the logical connectives but also upon whether its components parts are supported by the available empirical evidence, the calculus is described as an applied or interpreted deductive system.

^{5/} The concepts of 'pure' and 'applied' deductive systems are discussed in greater detail in: R.B. Braithwaite, Scientific Explanation, op. cit., Chapter 2.

2. Interpreted Systems and Empirical Science

A deductive system can function as a theory in empirical science only if some of the extra logical terms within the system are given an empirical interpretation. Such an interpretation can be provided by a set of sentences which relate certain terms in the deductive system to specific observational terms. When a deductive system has been sufficiently interpreted such that some of its deduced hypotheses can be directly submitted to empirical test, then such a system can be classified as a scientific theory. The theory itself may, on being subjected to test, be refuted or it may survive each test to which it is put. The point to note, however, is that until a theory contains some hypotheses which can be directly submitted to empirical test, the theory is not sufficiently interpreted to be classified as a scientific theory.

Consider for a moment the pure deductive system represented in the earlier example. In order to turn this system into a testable theory suitable interpretations need to be provided for the symbols ' Σ ' and ' Λ '. If one chooses the language of physics as the basis upon which to interpret these signs, then ' Σ ', the only undefined concept, could be interpreted in terms of the intersection of two fine hairs or in terms of the intersection of two specific rays of light. In either event, once a suitable interpretation for ' Σ ' is selected one then has an interpretation for ' Λ ', since ' Λ ' is defined solely in terms of ' Σ '. In this particular simple example a suitable interpretation of ' Σ '

is sufficient to provide a physical interpretation of the entire deductive system. Accordingly, it is now possible to inquire into the empirical relevance of this system. Indeed, one can directly investigate whether as a matter of empirical fact 'there are at least three points;' and whether 'there are at least three lines.' In the same manner it is possible to submit, in this particular example, the remainder of the deducible hypotheses to a direct confrontation with empirical evidence.

Clearly I could have chosen a slightly different example. I could have selected for the postulates the axioms for what is usually called Euclidean Geometry. If this deductive system is provided with a suitable physical interpretation one could again subject a number of the inferred hypotheses of this theory to empirical test. For once one has ascribed a physical interpretation to the terms 'angle' and 'degree' one can then test the hypothesis that 'the sum of the three angles in any triangle is one hundred and eighty degrees.'

In performing such a test on an applied deductive system it may well happen that the evidence confutes the hypothesis. If this occurs then it readily follows that under this particular interpretation, for example, it is a matter of empirical fact that 'the sum of the three angles in any triangle is not one hundred and eighty degrees.' While this may well appear as an obvious conclusion it must not be forgotten that within the pure deductive system this hypothesis remains a valid statement. What has happened is that by placing a particular physical interpretation on the system part of this theory has been shown to be empirically false with respect to that specific interpretation.

Now consider again, for a moment, the pure deductive system. Also suppose that I have a theory which is a physical interpretation of this system. Suppose further that both the deductive system and its physical interpretation are well developed, and that the theory has been submitted to and survived a number of empirical tests. Then, within the context of its physical interpretation the theory would be a corroborated theory of empirical science. I now devise a second physical interpretation for the basic deductive system where I ascribe new empirical meanings to the terms of the original system. Given the new interpretation I can now develop this new theory by taking advantage of all the hypotheses (or theorems) that have already been proved in the original deductive system. Strictly by placing the new interpretations upon the symbols contained in the original hypotheses it is possible to write down at once many of the hypotheses of my new theory. Whether these hypotheses will be corroborated or refuted is a matter for experiment and test. But the point to note is that because the second theory is a new interpretation of a basic system, one which has already been successfully interpreted into a tested theory, the theorems of the basic system can be immediately accepted as part of the new theory.

For example, if the basic system is a pure deductive system of geometry, and if the postulates of the new theory are particular hypotheses of physics, then if it can be shown that these postulates are simply a specific interpretation of the pure geometry, all the theorems of this geometry become, once suitably interpreted,

hypotheses of the new physical theory.^{6/}

3. Testing the Interpreted Systems

Once we have an interpreted deductive system or theory we are then in a position to inquire how and under what conditions we are able to submit this theory to empirical test.^{7/} The first and most obvious requirement is that if none of the theory's hypotheses (these include all hypotheses that are deducible within the confines of the deductive system) are stated solely in terms of observables, then none of the hypotheses can be directly confuted by empirical test.

The significance of this rather obvious requirement is due to the way in which meaning is given to propositions (hypotheses) stated in their normal, conditional form (see). As was noted above a conditional statement can only be shown to be false if there is evidence affirming the truth of the antecedent clause. If it is not possible to observe whether 'P' is true, then to find evidence supporting the whole statement 'P → Q' does not allow us to infer anything about the truth value

^{6/}A penetrating discussion of this point is to be found in: C.G. Hempel, "Geometry and Empirical Science," American Mathematical Monthly, Vol. 52, 1945. (Reprinted in: H. Feigl and W. Sellars, (eds) Readings in Philosophical Analysis, Appleton-Century-Crafts, New York, 1949, pp. 238-249.

^{7/}An excellent history of the criteria of empirical validity is to be found in: C.G. Hempel, "Problems and Changes in Empirical Criterion of Meaning," Revue Internationale de Philosophie, Vol. 11, 1950. (Reprinted in: L. Linsky, (ed), Semantics and the Philosophy of Language, University of Illinois Press, Urbana, 1952, pp. 163-185.

of the clause 'P'. Since the proposition 'P → Q' can have the value 'true' when both 'P' and 'Q' are false, it follows that we must have some knowledge concerning the empirical truth value of 'P' before it is pertinent to subject the hypothesis 'P → Q' to an empirical test.

If none of the theory's hypotheses are stated in terms of observables then the theory cannot provide us with a proposition within which one can determine the empirical truth value of the antecedent clause. Under this condition it is not possible to test for the empirical truth value of any of the theory's propositions. Consequently, the theory cannot be refuted by empirical test. If a theory only contains hypotheses which cannot be refuted then the theory itself is not saying anything about the world of empirical science. For once a theory makes an empirical claim then it is at the same time denying the opposite of that which it claims. If it is not possible to refute any of a theory's hypotheses then the theory is not denying anything. Accordingly, it is equally obvious that under these conditions the theory cannot be making any positive, empirical claims. Hence if a theory--that is, any of its hypotheses--cannot be refuted by empirical test then the theory cannot be considered a part of empirical science.

For a theory to be a part of empirical science at least one of its hypotheses must be stated in terms of observables. Assuming, for the moment, that we are considering such a theory. Let us now examine the different ways in which it can be submitted to empirical test. Since a theory consists of certain basic postulates, some definitions, some hypotheses and some interpretive rules we can imagine conducting our empirical tests upon the

postulates as well as the hypotheses. If the interpretive rules are such that the postulates themselves can be directly submitted to test then these tests can be performed independently from the theory for which they are the deductive base. If these postulates survive such tests then not only can they be considered as empirical hypotheses but they also serve as a strong empirical base for the resulting theory.

For example, consider an entire theory as consisting of a long sequence (conjunction) of conditional statements. Then the postulates are the initial statements in a number of sequences, where these series of statements are a part of the entire sequence of propositions representing the theory. Since the postulates are empirical hypotheses they provide the empirical truth value for the antecedent clause in each of the sequences of which they are the initial clause. In sequences where they are employed in other steps in the deductive process they again perform the function of imparting an empirical truth value to that part of the sequence. Consequently, a number of hypotheses can be identified for which the empirical truth value of the antecedent clause can be inferred. Manifestly, with such hypotheses, as long as the consequent refers to observables the hypothesis itself can be submitted to empirical test.

Since all theories do not contain basic postulates which are by themselves empirical hypotheses it is necessary to consider the conditions under which the remaining classes of theories can be submitted to empirical test. Clearly, if the postulates cannot be tested directly it must be possible to deduce some hypotheses which can be submitted to test.

Such hypotheses, as we have seen, must be stated in terms of observables. And if they are corroborated by such tests as they are exposed to then these hypotheses in turn provide the empirical support for the hypotheses from which they were inferred.

It should be noted, as well, that to be able to test a general hypothesis it must be possible to deduce from it a singular instance against which the experimental data are to be applied. One cannot test a general hypothesis or law by employing general data. The best that can be done is to repeatedly test the general law by subjecting further specific instances of it to experimental test.

An obvious example of this process can be found if by examining the general law which asserts that: every body near the earth that is freely falling towards the earth falls with an acceleration of 32 feet per second per second. To test the empirical validity of this law one cannot employ general data about every free falling body near the earth's surface. To test it at all one must deduce from the general law the singular statement that: a body starting from rest and freely falling towards the earth falls $16t^2$ feet in t seconds. By specifying some initial conditions--namely, that t shall have the value of one second,---the directly testable statement is inferred that: a body starting from rest and freely falling towards the earth falls 16 feet in one second. This, and numerous other such singular instances of the general law, can be subjected to the process of refutation by empirical test. And so long as the evidence does not confute these statements the general law is accepted as an empirical hypothesis which has yet to be disconfirmed.

4. Scientific Explanations and Predictions^{8/}

Once we have a testable theory of some particular class of phenomenon we are then in a position to inquire how to employ this theory to establish explanations of the occurrence of those events. To explain the occurrence of a specific event part or all of the theory is employed in the following manner:

The theory itself provides the hypotheses and delineates the initial conditions that must be taken into account if the event is to be explained. The underlying deductive system provides the rules of inference by which, from the conjunction of the initial conditions and the relevant hypotheses, we deduce the occurrence of the event in question. Hence, an explanation is established by deducing the occurrence of an event from the conjunction of the theory's hypotheses and a specific set of observable initial conditions.

^{8/} The term "scientific explanation" has no honorific connotations. It is used to distinguish the type of explanation discussed in this book from teleological, as well as other types of explanation found in discussions of the explanatory process. Accordingly throughout this essay the terms "scientific explanation" and "explanation" are used synonymously.

One further item concerns the type of explanatory process discussed in this section. The following pages are concerned with a brief description of what is called the deductive-nomological schema. Later in this essay probabilistic explanations and predictions will be considered. For a detailed exposition of the explanatory process see: C.G. Hempel and P. Oppenheim, "The Logic of Explanation," Philosophy of Science, Vol. 15, 1948, (Reprinted in H. Feigl and M. Brodbeck, (eds), Readings in the Philosophy of Science, Appleton-Century-Crafts, 1953, pp. 319-352.

We have already seen that for a theory to be a part of empirical science it must contain at least one hypothesis that can be submitted to empirical test. But before a theory can be employed to establish a scientific explanation it must contain at least one hypothesis that has survived a number of empirical tests. It is not sufficient that at least one hypothesis is refutable, it must have been demonstrated to be able to withstand empirical tests.

Given that the theory contains at least one general, empirical hypothesis the other main requirement for an explanation is that the statements describing the initial conditions be empirically true. If we are to deduce the occurrence of an event from the conjunction of a set of hypotheses and initial conditions then just as at least one hypothesis must have been submitted to empirical test so the initial conditions must also be empirically true if the explanation itself is to be empirically true.

The same conditions, of course, must be met if we are to establish a scientific prediction of the occurrence of a particular event. These requirements must be met for the same reasons as were put forward when the process was described by which a hypothesis or theory is submitted to empirical test. If the initial conditions are not known to be empirically true, and if at least one hypothesis within the theory has not survived a number of empirical tests, then to employ such a theory to predict the occurrence of an event is an empirically meaningless affair.

Frequently, the predictions of a theory are employed as a way of submitting the theory to empirical test. If by this process a correct

prediction is produced then before it is scientifically meaningful as evidence of the theory's empirical validity, it must be shown that the theory contains at least one empirical hypothesis and that the initial conditions were empirically true. Otherwise we are in the position which can be represented by the case where we are dealing with the conditional statement ' $\underline{P} \rightarrow \underline{Q}$ ' and where we do not know the empirical truth value of ' $\underline{P}$ '. If ' $\underline{P}$ ' is false then ' $\underline{P} \rightarrow \underline{Q}$ ' is true whether ' $\underline{Q}$ ' is true or not. Hence, solely by correctly predicting and observing the occurrence of ' $\underline{Q}$ ' we have not learned anything about the empirical truth value either of ' $\underline{P} \rightarrow \underline{Q}$ ' or ' $\underline{P}$ '. But if from empirical observation ' $\underline{P}$ ' is true then to correctly predict and observe ' $\underline{Q}$ ' permits us to corroborate ' $\underline{P} \rightarrow \underline{Q}$ '. Hence, a prediction has the same logical form as an explanation and must meet the same requirements if it is to be classified as a part of empirical science.

Summary

Having examined the logical structure of a theory, the conditions under which it can be subjected to empirical test, and the manner in which it can be employed to establish scientific explanations and predictions it will aid the analysis in the following pages if author and reader are both quite clear on what it is that a theory is expected to do. In particular, what characteristics or properties are expected of a theory of economics? A comprehensive answer to this question is given by Quine^{9/}, and by a direct interpretation the following statement

^{9/}Williard Van Orman Quine, From a Logical Point of View, Harvard University Press, 1953, pp. 53-54.

describes what we might expect from a theory in microeconomics.

One begins by describing the desired class of significant sequences of observed behavior to which the theory refers, as the class K. Class K is itself the culmination of four classes, H, I, J, K, of observable statements of economic behavior. These classes are of increasing size and are described as follows: H is the class of observed sequences of behavior, excluding any which are ruled inappropriate in the sense of being beyond the scope or domain of the theory. For example, if the theory is concerned with the production and distribution process of the firm, the observed behavior of consumers might be ruled as being beyond the domain of the theory. Similarly, statements of observed behavior belonging to other subject matter such as physiology or biology, would also be ruled as inappropriate to H. I is the class of all such observed sequences of economic behavior and all that will ever be professionally observed, excluding again those which are ruled inappropriate. Hence, for class I the theory of the production and distribution process of a firm would include all professionally observed sequences of behavior relating to the production and distribution process of the firm. J is the class of all observable sequences of economic behavior ever occurring, now or in the past or the future, whether professionally observed or not, excluding once again, only those sequences which are ruled to be inappropriate. K, finally, is the infinite class of all those sequences of behavior, excluding the inappropriate ones as usual, which could be observed. Hence, K is the class which the theorist would like to approximate in his formal reconstruction, where K is more inclusive than J, notwithstanding H and I. From this description it can be seen

that the class of statements contained in H constitutes at best the growing record. Class J on the other hand includes statements that go beyond any record even though these statements still have a certain common sense reality. However, very little can be said about the reality of the statements included in K because of the word 'could'.

A theory that is constructed by making observations within H can and, as a rule, is tested by applying it to other sequences of observed behavior in H. If the theory survives in H and is not confuted by any part of it, then it is within the class of I that disconfirmation or continued corroboration must come. Although the theory may be presumed to hold in J and K, science is restricted to testing its theories against observable sequences of behavior. Consequently, a theory can only be subjected to test within H and J. As a result, no matter what our feelings are toward the "eternal truths" embodied in any particular theory, the domain of observable events over which we may speak of a theory holding--i.e. not having been disconfirmed--is in practice delimited by the classes H and I.

As scientists we are committed to the working hypothesis that although our theories may be false in J we can only talk about their empirical validity within H and I. Thus, if a theory is to be empirically testable, and if it is to explain and predict the occurrence of specific events, then we can conceive of its function as that of being able, in principle, to generate all of the relevant, observable sequences of behavior included in H and I. It is in carrying out this investigation--namely, of seeing whether the sequences of behavior generated by the theory conform with those contained in H and I--that we in fact submit the theory to test and determine whether it can be accepted as empirically confirmed.

PART TWO

Mathematical Foundations

of

Classical Analysis

Chapter 3

CLASSICAL FOUNDATIONS OF ECONOMIC ANALYSIS^{1/}

In order to be able to understand the way in which economic theory has been and currently still is developed it is necessary to examine the basic deductive system from which the theory is derived, as well as the empirical interpretation that is placed upon the terms once the theory is constructed. This chapter is concerned with an investigation of the former--i.e. the mathematical basis of microeconomic theory--leaving for later chapters the task of examining the rules by which these theories are empirically interpreted.

To help focus attention on the significant properties of the mathematical foundations consider the following simple example of the way in which an answer is produced to a specific economic question. The problem is to determine what the effect will be on the output of a firm if a tax on output is imposed. To answer this question we need to know the relation between a tax on output and the firm's output itself. That is to say, if such a tax is levied will the firm increase or decrease its output? Or if such a tax is already in effect and is further increased will output increase or decrease. The answer to these questions is provided in the following way:

First consider a firm for which the demand curve for its goods and

^{1/} This chapter is primarily based upon P.A. Samuelson's classic book, Foundations of Economic Analysis, Harvard University Press, 1947. Further, in order to facilitate reference to this excellent analysis of the mathematical foundations, the notation employed in this chapter is kept the same wherever possible.

services is known. For simplicity take the case where the firm produces only one item. Then our knowledge of the demand curve will tell us the quantity of items the firm will produce at various market prices. To proceed with the analysis we also need to know the relation between the total production cost for the firm and its output--the production cost schedule. Given that these two main items are known we can then state that the total profit for the firm for any particular price for its product is the difference between its total revenue and its total cost:

If x represents the quantity sold.

$p(x)$ represents the market price for this quantity

$C(x)$ represents the lowest total product cost at which each output is produced

Π represents total profit.

then

$$\Pi = x p(x) - C(x)$$

Now if a tax of t dollars is imposed on each unit of output x , then the total tax payment is given by tx . Thus, after such a tax is imposed the firm's total profit is:

$$\Pi = x p(x) - C(x) - tx \quad (3.1)$$

Such, then, is the general form of the effect of a tax on output upon the profit of the firm. But, before this statement asserts anything about the firm's behavior in response to this tax it must be possible to write down the relations assumed under the terms $x p(x)$ and $C(x)$. Further we must be able to produce the requisite initial conditions, i.e. in this case a specific tax rate of, say, t^0 dollars per unit. Once the tax rate is given then the output of the firm at this tax rate, x^0 , can be represented by:

$$x^0 = g(t^0) \tag{3.2}$$

where the relation $g(t^0)$ represents that specific set of parameter values in the equations representing $x_p(x)$ and $C(x)$ which will yield an output of x^0 for a tax per unit of (t^0) .

For any particular collection of equations representing $x_p(x)$ and $C(x)$ there are a variety of different sets of parameter values which will yield an output x^0 for an initial condition of t^0 . Each of these different sets of parameter values represents a solution to the total system of equations. Since there can be as many possible solutions the (sets of parameter values) as there are equations in the theory some restrictions must be introduced if we are to have a unique solution to the problem. If such restrictions are not imposed, then each real solution provides a separate answer each of which is as valid as any other. Accordingly, without a rule for selecting among solutions we would be unable to arrive at a single answer to the original question.

To generate a unique solution it is assumed that the firm will select for the given tax rate t^0 the particular output which will maximize its profit (net revenue). By introducing this assumption the possible set of different solutions are restricted to exactly that one set which will satisfy this maximum or equilibrium restriction. The answer is arrived at by applying the conditions for this equilibrium to the theory of the firm's behavior represented by (3.1).

The equilibrium conditions for a regular maximum of profit with respect to the output under a specific tax rate are

$$\frac{\partial \Pi(x, t)}{\partial x} = 0 \quad (3.3)$$

$$\frac{\partial^2 \Pi(x, t)}{\partial x^2} < 0 \quad (3.4)$$

The first condition (3.3) states that for the firm to be maximizing net reserve it must be at a point such that the slope of the function depicting profit against output is equal to zero. While condition (3.4) insures that we do have a maximum by excluding the cases of being at a minimum or a saddle point.

Applying condition (3.3)^{2/} to the original equation (3.1) the first restriction is written as

$$\frac{\partial}{\partial x} [xp(x) - C(x) - tx] = 0$$

or

$$\frac{\partial}{\partial x} [\dot{x}p(x) - C(x)] - t = 0$$

$$\text{Hence} \quad t = \frac{\partial}{\partial x} [xp(x) - C(x)] \quad (3.5)$$

By solving equation (3.5) we determine its roots. By taking the second differential of (3.5) with respect to $\underline{x}$ we can ascertain whether the solutions developed for (3.5) represent the position of a maximum or not. If condition (3.4) is satisfied for all relevant values of $\underline{x}$ and $\underline{t}$, then we have determined the equilibrium quantity $\underline{x}$ that corresponds to the tax rate $\underline{t}$.

In equation (3.2) it was noted that for a specific set of parameter

^{2/}By applying conditions (3.3) and (3.4) to equation (3.1) we are implicitly assuming that all the function represented in (3.1) contain continuous first and second derivatives everywhere over the relevant domain.

values in the functions $x_p(x)$ and $C(x)$ a tax rate t yields an output x^0 . In equation (3.5) the same result is described in a more general form. In (3.5) it is clear that for each value of t there will be a set of parameter values for the functions on the right hand side, and that this set of values will be the equilibrium solution for those specific values of t .

Once the firm's output for each tax rate is determined (as long as the firm continues to maximize its profit) the next step is to discover how the values of the equilibrium output behave with respect to changes in the tax rate. Presumably, this question can be answered by solving equation (3.5) and working out numerical answers for different values of t . To perform these calculations the relations represented by $x_p(x)$ and $C(x)$ must be specified for a particular firm. Further, these relations must be stated in a form that is suitable to analytic or numerical analysis. Given that the requisite knowledge is available explicit numerical solutions can be calculated and the results plotted. Accordingly, one can graphically determine the rate at which the equilibrium output alters with respect to changes in the tax.

However, from our knowledge of the equilibrium conditions we can also analytically derive some conclusions about the rate of change of output. Since we have already deduced that the equilibrium output for each value of the tax rate is given by

$$t = \frac{\partial}{\partial x} [x_p(x) - C(x)] \quad (3.5)$$

we can differentiate (3.5) with respect to t to arrive at the following rate of change as follows:

$$\frac{\partial}{\partial t^0} [t] = \frac{\partial}{\partial t^0} \left\{ \frac{\partial}{\partial x} [x^0 p(x^0) - C(x^0)] \right\}$$

where $\underline{x}^0$ and $\underline{t}^0$ denotes that we are differentiating with respect to equilibrium values.

Carrying out the differentiation this equation is reduced to

$$1 = \left(\frac{\partial x}{\partial t} \right)^0 \left\{ \frac{\partial^2}{\partial x^2} [x^0 p(x^0) - C(x)] \right\} \quad (3.6)$$

From (3.6) it follows that the equilibrium rate of change of output with respect to $\underline{t}$ is given by

$$\left(\frac{\partial x}{\partial t} \right)^0 = \frac{1}{\frac{\partial^2}{\partial x^2} [x^0 p(x^0) - C(x^0)]}$$

But we also know from the previous analysis of equation (3.5) that for all relevant values of $\underline{x}$ and $\underline{t}$

$$\frac{\partial^2}{\partial x^2} [x^0 p(x^0) - C(x^0)] < 0 \quad \underline{3/}$$

Hence it follows that

$$\left(\frac{\partial x}{\partial t} \right)^0 < 0$$

which states that as long as the firm is maximizing its net revenue both before and after the tax on output is applied then as the tax rate is increased (decreased) the output will decrease (increase). Without specifying in complete detail the functions represented by $\underline{x}p(x)$ and $\underline{C}(x)$ it is not possible to determine the specific rate at which output will

^{3/} This being a sufficient condition to assure a relative maximum, see relation (3.4).

increase or decrease. But without employing such detail we can determine, under the equilibrium restriction, the direction of this rate of change.

In this example two basic analytic techniques are employed that characterize a large part of the mathematical foundations of microeconomics. But in order to derive a clearer understanding of exactly what they entail for economic analysis we need to examine them in additional detail.

1. The Technique of Equilibrium Solutions

In the example of the tax on a firm's output it was assumed that we knew the demand and production cost schedules for a firm. While these relations were represented by two terms, $x_p(x)$ and $C(x)$, this system of functions were neither specified nor examined in any detail. To do so let us take the general case where the economic system, including the demand and production cost schedules, can be described by n variables $(x_1, x_2, \dots, x_n)$ and m parameters $(\alpha_1, \alpha_2, \dots, \alpha_m)$ where $m \begin{matrix} < \\ > \end{matrix} n$. It is assumed that these n variables and m parameters are contained in n independent and consistent functional relations.

In the tax example we supposed that we knew the demand and production cost schedules for the firm. Accordingly we were supposing that we could write down in complete detail the specific equations, variables and parameters which pertained to that firm.

In general terms the total system of functional relations can be written in the following way:

$$\begin{aligned}
 f^1(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) &= 0 \\
 f^2(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) &= 0 \\
 &\vdots \\
 f^n(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) &= 0
 \end{aligned} \tag{3.7}$$

where each relation is represented as being a function of all n variables and m parameters. Of course, in any particular case the values of many of the parameters for each relation would be equal to zero. But when considering the general case it is useful to express the system as noted in (3.7).

Given n independent and consistent functional relations there are in general n possible sets of solutions. In other words, it is possible that there are n different sets of values of the parameters $(\alpha_1, \alpha_2, \dots, \alpha_m)$ which correspond to n different values of the variables $(x_1, x_2, \dots, x_n)$, each set of values representing one solution to the functional relations.

In the tax example the possible set of solutions were reduced to a single one by imposing the conditions for equilibrium (3.3) and (3.4) upon the system. The addition of these conditions determined a unique value for the variable x which corresponded to a given value of t. These two values are labelled x^o and t^o. For the general case we can employ the same notation and represent the equilibrium solution for the system by that set of values for the variables $(x_1^o, x_2^o, \dots, x_n^o)$ which correspond to the given values of the parameters $(\alpha_1^o, \alpha_2^o, \dots, \alpha_m^o)$. The equilibrium values of x_i^o are thus a function of some specific parameter values $(\alpha_1^o, \alpha_2^o, \dots, \alpha_m^o)$. This relation can be expressed as:

$$x_i^o = g^i(\alpha_1^o, \alpha_2^o, \dots, \alpha_m^o) \quad (i=1, 2, \dots, n) \tag{3.8}$$

which clearly corresponds to relation (3.2).

In order to discover the equilibrium values $\underline{x}_i^0$ in the general case the procedure is to impose a set of initial conditions (choose a set of parameter values) and to constrain the possible set of solutions by employing the first and second order conditions: For the general case these constraints are represented by:

$$\frac{\partial}{\partial x_i} f^i(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) = 0 \quad (3.9)$$

$$\frac{\partial^2}{\partial x_i^2} f^i(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) < 0 \quad (3.10)$$

If (3.10) is satisfied for the relevant parameter values then the solution to the system of relations represented by (3.9) is the unique set given by (3.8), i.e.

$$x_i^0 = g^i(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0) \quad (i=1, 2, \dots, n)$$

Consequently, for any given case if we are able to completely specify the relations contained in (3.7), then by selecting the initial conditions and by imposing (3.9) and (3.10) we are able to arrive at the unique solution represented by (3.8). Further, by solving for the relations expressed in (3.8) we can then determine the equilibrium values for all the variables of the system that correspond to this particular set of parameter values.

In the problem on the effect of a tax on the output of a firm the equilibrium solution for the tax rate $\underline{t}$ is given by:

$$t = \frac{\partial}{\partial x} [xp(x) - C(x)]$$

For an initial value of $\underline{t}^0$ there corresponds an unique value of $\underline{x}$ given by $\underline{x}^0$. This value can be precisely computed if the relations represented by $\underline{xp(x)}$ and $\underline{C(x)}$ are known. In this specific case it should be noted that the equilibrium

conditions correspond to the behavioral postulate that firms behave so as to maximize their net revenue. Accordingly, given this postulate we are able to infer that the firm in question is operating at a point of maximum net revenue which in turn permits the deduction of the relation between the tax rate and output as noted above.

Returning to the general case it is clear that if the economic system under investigation can be represented by n independent and consistent functional relations consisting of n variables and m parameters, then an unique solution to this system can always be determined if and only if the conditions for equilibrium representing some behavioral postulates about the system are imposed.

2. Comparative Statics and the Displacement of Equilibrium

Once we know the relations governing a specific economic system we can, with the aid of a maximizing or similar equilibrium postulate, determine its equilibrium solution. In most practical applications, however, a variety of difficulties are encountered. The first and most obvious obstacle is that it is not always possible to state the precise set of relations of which the system under investigation is composed. While we may be able to note its general form as well as some of the variables that must be included, we are frequently unable to specify for a particular case the empirical relations that constitute the system. A second difficulty is raised by the parameter values themselves. Not all the variables in these economic relations refer to items that can be directly observed. As a result, a number of parameters do not refer to observational items. But to generate

a specific, equilibrium solution the initial values of the parameters must be specified. Consequently, within any particular application it may not be possible to derive a solution which is stated entirely in terms of observables.

If specific equilibrium solutions could always be generated for each economic application, then it would always be possible to plot the way in which a variable of interest altered given certain changes in the initial conditions. In the tax example this would mean that if one were able to completely specify the solution of $t = \frac{\partial}{\partial x} [xp(x) - C(x)]$, one would then be able to plot the different values of x that would correspond to a variety of assumed tax rates. Once this was accomplished, and as long as the system of relations continued to represent the economic system in question, then one would be in a position to state, for example, exactly how the equilibrium output of this firm would respond to alterations in such a tax rate.

Unfortunately, we are not always going to be able to derive specific equilibrium solutions for which the results may be plotted by varying the initial conditions. At the same time, for a variety of reasons, we still would like to know something about the responses of the particular economic system to selected changes in the initial conditions. As a result an additional technique is required, known as comparative statics, which allows us to determine the directional change of individual variables in response to a selected change in the initial conditions.

The method of comparative statics proceeds by subjecting the equilibrium solution to shifts in its parameter values. These shifts are then employed to

determine the slope or directional change of selected variables. The entire analysis is conducted under equilibrium restrictions so that these conditions can be employed to help derive the unique solution to the new system. More specifically, suppose that we wish to determine the rates of change of some or all of the variables within the economic system with respect to a change in the initial value of, say, α_1 . To effect such an analysis we proceed in the following way:

We begin, of course, with the current equilibrium solution to the system. That is to say, we already have the initial conditions given by the parameter values $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ with the corresponding set of variables $(x_1^0, x_2^0, \dots, x_n^0)$. At the same time we have the system of relations at their equilibrium values given by

$$f^i(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0) = 0 \quad (i=1, 2, \dots, n)$$

as well as the solution to this system given by

$$x_i^0 = g^i(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0) \quad (i=1, 2, \dots, n)$$

We now want to determine the rate of change of the variables with respect to α_1 . To accomplish this we take the first derivative of the variables in the relations $f^i(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ with respect to α_1 , while at the same time holding the values of all the other parameters constant. By taking the partial derivative of each relation with respect to α_1 and holding all other parameter values constant we generate a system of n relations as follows:

$$\begin{aligned}
 f_{x_1}^1 \left(\frac{\partial x_1}{\partial \alpha_1} \right)^0 + f_{x_2}^1 \left(\frac{\partial x_2}{\partial \alpha_1} \right)^0 + \dots + f_{x_n}^1 \left(\frac{\partial x_n}{\partial \alpha_1} \right)^0 &= -f_{\alpha_1}^1 \\
 f_{x_1}^2 \left(\frac{\partial x_1}{\partial \alpha_1} \right)^0 + f_{x_2}^2 \left(\frac{\partial x_2}{\partial \alpha_1} \right)^0 + \dots + f_{x_n}^2 \left(\frac{\partial x_n}{\partial \alpha_1} \right)^0 &= -f_{\alpha_1}^2 \\
 \vdots & \\
 f_{x_1}^n \left(\frac{\partial x_1}{\partial \alpha_1} \right)^0 + f_{x_2}^n \left(\frac{\partial x_2}{\partial \alpha_1} \right)^0 + \dots + f_{x_n}^n \left(\frac{\partial x_n}{\partial \alpha_1} \right)^0 &= -f_{\alpha_1}^n
 \end{aligned} \tag{3.11}$$

where the symbols

$$f_{x_j}^i = \frac{\partial f^i}{\partial x_j} (x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$$

and

$$f_{\alpha_1}^i = \frac{\partial f^i}{\partial \alpha_1} (x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$$

and where within each relation all the remaining variables and parameters are kept constant.

These relations while formidable in appearance correspond directly to the single relation given by (3.6). In that case the term

$f_{x_j}^i = f_x^1 = \frac{\partial}{\partial x} [x^0 p(x) - C(x)]$, while the term $f_{\alpha_j}^i = f_{t_0}^1 = \frac{\partial}{\partial t_0} [t] = 1$. After performing the partial differentiation of $\underline{x}$ with respect to $\underline{t}^0$ we were left

with the expression

$$\left(\frac{\partial x}{\partial t} \right)^0 \left\{ \frac{\partial^2}{\partial x^2} [x^0 p(x^0) - C(x^0)] \right\} = 1 \quad .$$

To solve this equation in terms of the desired rate of change, $\left(\frac{\partial x}{\partial t} \right)^0$, the second order condition for equilibrium was employed to show that the value of

$$\frac{\partial^2}{\partial x^2} [x^0 p(x^0) - C(x^0)] < 0 \quad .$$

Thus, we could immediately conclude that the value of $\left(\frac{\partial x}{\partial t} \right)^0$ was also less than zero.

In the general case we proceed towards a solution with respect to the desired rates of change $\left(\frac{\partial x_i}{\partial \alpha_1}\right)^o$ in a similar way.

If we examine the relations expressed in (3.11) more closely we see that the $f_{\alpha_1}^i$ are all constant terms and that the $f_{x_j}^i$ represent the coefficients of the $\underline{n}$ variables $\left(\frac{\partial x_i}{\partial \alpha_1}\right)^o$. Accordingly, (3.11) is a system of $\underline{n}$ linear equations in $\underline{n}$ unknowns.

The solution for such a system of linear equations for the non-singular case is readily represented in matrix terms:^{4/}

$$\left(\frac{\partial x_s}{\partial \alpha_1}\right)^o = -\frac{|A_s|}{|A|} = -\frac{\sum_{i=1}^n f_{\alpha_1}^i A_{is}}{|A|} \quad (s=1,2,\dots,n) \quad (3.12)$$

$$\text{where } |A| = \begin{vmatrix} f_{x_1}^1 & f_{x_2}^1 & \dots & f_{x_n}^1 \\ f_{x_1}^2 & f_{x_2}^2 & \dots & f_{x_n}^2 \\ \vdots & \vdots & & \vdots \\ f_{x_1}^n & f_{x_2}^n & \dots & f_{x_n}^n \end{vmatrix}$$

and where A_{is} represents the cofactor of the element of the i th row and the s th column.

In order to solve for the exact values of $\left(\frac{\partial x_s}{\partial \alpha_1}\right)^o$ it is necessary to know all the values of all the $f_{x_j}^i$ terms appearing in (3.11). But all we primarily interested in knowing is the algebraic sign of the $\left(\frac{\partial x_s}{\partial \alpha_1}\right)^o$ terms. Consequently, although we do not need to know the exact values of the $f_{x_j}^i$

^{4/}This procedure is examined in more detail in Appendix A.

terms, we do need to know their algebraic sign.

To evaluate (3.12) in terms of its signs it must be possible to determine the sign of $|A|$ and A_{is} .

The value of $|A|$ is given by the product of $\underline{n}!$ terms, each of which is the product of $\underline{n}$ elements. If we are to be able to deduce the sign of $|A|$ we clearly need to know the signs of the $\underline{n}!$ terms. Further to know, unambiguously, the signs of the $\underline{n}!$ terms we need to know the signs of the individual $\underline{n}$ elements which comprise these terms. Since these individual elements are the signs of the f_x^i 's we seem to be caught in an unpleasant position. For by this analysis it appears that we must know the direction of the rates of change before we are able to solve for them. Clearly, this does not constitute a solution to the problem. And in order to effect a solution a slightly different direction approach is required.

A solution is achieved by placing a number of restrictions upon the way in which the signs of the unknowns, $\left(\frac{\partial x}{\partial \alpha_1}\right)^0$, are allowed to shift with respect to changes in selected parameter values. The first restriction requires that cases only be considered where one parameter is allowed to shift at one time. The second condition requires that this change in the parameter may only alter one of the equilibrium relations as are represented in (3.11). By requiring that a change in the $\underline{i}$ th parameter must leave all others but the $\underline{i}$ th equation unchanged, the rate of change of the $\underline{i}$ th variable to the remaining parameters must be equal to zero. Under these conditions it is possible to arrive at a criterion with which we can ascertain the direction of the rate of change of the $\underline{i}$ th variable in response to a shift in the $\underline{i}$ th parameter--i.e. the sign of $\left(\frac{\partial x_i}{\partial \alpha_i}\right)^0$ is determinable.

The criterion is derived from the first and second order conditions for an equilibrium and is given by the following statement:^{5/}

$$f_{x_i \alpha_i} \left(\frac{\partial x_i}{\partial \alpha_i} \right)^0 > 0 \quad (3.13)$$

This criterion states that under equilibrium conditions the rate of change of the ith variable with respect to a shift in the ith parameter is of the same sign as $f_{x_i \alpha_i}$, where

$$f_{x_i \alpha_i} = \left(\frac{\partial}{\partial x_i} \right) \left(\frac{\partial}{\partial \alpha_i} \right) f(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_n^0)$$

If the shift in the equilibrium equation is in the same direction as an increase in x_i , then $f_{x_i \alpha_i}$ is positive. Under this condition $\left(\frac{\partial x_i}{\partial \alpha_i} \right)^0$ is also clearly positive. Similarly if the shift in the equilibrium equation is in the opposite direction to an increase in x_i , then both $f_{x_i \alpha_i}$ and $\left(\frac{\partial x_i}{\partial \alpha_i} \right)^0$ are negative.

To illustrate the application of this criterion let us reconsider the example discussed earlier in this chapter. Once again the object is to discover the direction of the rate of change of the equilibrium output of a firm with respect to a shift in the tax rate. We have already seen that the net profit of the firm with a tax imposed upon its output is given by (3.1) or:

$$\Pi = xp(x) - C(x) - tx$$

Now if the firm is operating so as to maximize its net revenue than the equilibrium criterion states that the direction of the rate of change of

^{5/}The steps of the proof leading up to the statement of this criterion are given in Math Appendix.

output, $\underline{x}$, with respect to a shift in $\underline{t}$ is determined by (3.13). In this case the criterion is expressed by

$$f_{xt} \left(\frac{\partial x}{\partial t} \right)^0 > 0 \quad (3.14)$$

where

$$f_{xt} = \left(\frac{\partial}{\partial x} \right) \left(\frac{\partial}{\partial t} \right) \Pi = \left(\frac{\partial}{\partial x} \right) \left(\frac{\partial}{\partial t} \right) [xp(x) - C(x) - t(x)] .$$

Taking the two partial differentials of Π with respect to $\underline{x}$ and $\underline{t}$ we have in successive stages: first

$$\frac{\partial}{\partial t} \left[\frac{\partial \Pi}{\partial x} \right] = \frac{\partial}{\partial t} \left[\frac{\partial xp(x)}{\partial x} - \frac{\partial C(x)}{\partial x} - t \right]$$

then $\frac{\partial \Pi}{\partial xt} = -1$.

Substituting the value of f_{xt} into (3.14) we have

$$(-1) \left(\frac{\partial x}{\partial t} \right)^0 > 0 .$$

From which it follows that $\left(\frac{\partial x}{\partial t} \right)^0 < 0$, which states that the equilibrium output of the firm will move in the opposite direction to a shift in the tax rate.

Summary

The purpose of this chapter has been to examine the basic deductive system that underlies microeconomic theory. While one particular economic example was employed to briefly illustrate the application of this deductive framework to a specific problem, the principal concern has been to inquire into the basic components of this deductive apparatus. As a result of this investigation it is clear that the basic components of the mathematical foundations are as follows:

(1) If we are able to represent the economic system under investigation by n independent and consistent relations consisting of n variables or unknowns, then a unique solution to the values of these variables can be determined, only if we impose the first and second order conditions for equilibrium upon the system. By the imposition of these equilibrium conditions we also require that the n independent and consistent relations have continuous first and second derivatives everywhere over the relevant domain. (As a matter of fact certain analytic techniques allow us to be able to include relations which have selected types of discontinuities,^{6/} but aside from these specific exceptions the relations must have continuous first and second order derivatives.)

(2) Given that we can derive the equilibrium solution to our economic system we are then in a position to determine the direction in which specific equilibrium variables will shift in response to certain changes in specific parameter values. To be able to determine these directional shifts in the variables we need to restrict our attention to the particular case where an alteration in a single parameter affects the value of only one equilibrium variable. In other words to find the direction of the rate of change of one equilibrium variable with respect to a change in one parameter, all other variables and parameters must be treated as constants.

Stated in terms of the analytical framework presented in the previous chapter the deductive system can be represented in the following way: The basic postulates are the behavioral assumptions which are represented by the two equilibrium constraints (3.9) and (3.10). The hypotheses of the system

^{6/}See P.A. Samuelson, op. cit., pp. 70-73.

are the set of n independent and consistent relations which are represented by (3.7). The logical rules of inference are those which pertain to the differential and integral calculus for the simple reason that this is the calculus that is employed to express and manipulate the system's hypotheses. As expressed, for example, by the three sets of relations (3.7), (3.9), and (3.10) the deductive system can be called pure. Such a system becomes empirical significant when some or all of its terms are related to observational terms. Accordingly, in order to begin the exploration of the way in which these basic components are employed in the statement and development of economic theory the next chapter is devoted to an examination of the microeconomic theory of market equilibrium.

Chapter 4

THE THEORY OF MARKET EQUILIBRIUM

The theory of market equilibrium is a theory of the behavior of firms and consumers in the market place. The theory itself is stated in mathematical form and as a result the behavior of firms and consumers is represented by mathematical relations. Whether these relations can be sufficiently interpreted such that they themselves, or consequences deducible from them, can be confuted by empirical test is a question to be examined in later chapters. For the moment it is the process by which such a theory is developed that is of primary interest. Accordingly, this chapter is concerned with an examination of the specific application of the economic deductive system which leads to the microeconomic theory of market equilibrium.

The theory to be examined refers to a perfectly competitive commodity market. Although there are a number of theories which refer to specific, different types of markets, e.g. monopoly, monopsony, duopoly, oligopoly, to mention but a few of the possible cases, each is developed by a similar application of the rules and constraints of the economic deductive system to a specific set of initial postulates. Hence, the selection of the perfectly competitive market for examination in no way restricts the scope or relevance of the analysis of the process by which a theory of market equilibrium is developed.

The object of a theory of market equilibrium is to describe the mechanisms which determine the quantities bought and sold, as well as the prices at which these transactions take place. The market consists of consumers and firms, and the combination of all their purchases and sales constitutes the total volume of transactions. While the consumer may be the

buyer and the firm the seller in a market for one particular kind or class of goods, e.g. automobiles, the firm is the buyer and the consumer is the seller in the market for the consumer's labor. In a number of markets firms are both buyers and sellers, such as in the case of mining companies that produce and distribute various ores to manufacturing companies who purchase the ores. In all of these cases it is the interaction between buyer and seller that is of principal interest and to which the market theory is primarily addressed.

1. Conditions and Postulates of the Theory

A commodity market is classified as perfectly competitive if the following conditions are satisfied: (1) the firms operating within this particular market all produce a specific type of commodity such that the product of one firm is indistinguishable to the consumer from that of any of the other firms; (2) the consumers of this market are such that they all appear identical to the sellers so that there are no special advantages to be gained or losses to be incurred by selling to one consumer rather than another; (3) the number of consumers and firms is sufficiently large so that the purchases or sales of each specific unit is small in relation to the total volume of transactions within the market; (4) information on prices and quantities offered within the market is such that both consumers and firms have perfect knowledge of the current prices and bids; (5) there is no restriction on participating in the market--i.e. both consumers and firms are free to enter and leave the market at any time.

The first condition requires all firms within the market to be producing a homogeneous product. If brand names, special promotional schemes, trademarks, and other means for identifying products exist then the market does not satisfy this condition. The object of the condition is to identify a market in which consumers have no criterion other than price with which to distinguish the product of one firm against that of another. The second condition is a complement to the first in that it requires the consumers of this market to be indistinguishable from one another with respect to the sellers. As a result firms within this market have no criterion other than price by which they decide to sell their produce to one consumer rather than another.

The third condition requires that the number of consumers and firms be sufficiently large so that the purchases or sales of any one individual are not large enough to significantly affect the market price. Accordingly, both consumers and producers perceive the prevailing price for a particular commodity as the only one available and make their decisions to buy or sell on the basis of this price alone.

The fourth and fifth conditions ensure that both producers and consumers are fully aware of the current prices and bids for the commodities in question. If this information is such that the consumer or the producer decides to withdraw from the market then there is no restriction associated with this decision. As a result neither consumers nor producers can buy or sell except at the prevailing price, and the market is determined to be perfectly competitive if these conditions are satisfied by both buyers and sellers.

The first main hypothesis of the theory is that a consumer's demand for a specific product in the market is a function of the price of the product, the prices of all other products within the market, and his income. Consider the $\underline{i}$ th consumer and the $\underline{j}$ th commodity. Then this hypothesis can be stated as follows:

$$\underline{D}_{ij} = \underline{D}_{ij}(p_1, p_2, \dots, p_m, y_i) \quad (i=1, 2, \dots, n) \quad (j=1, 2, \dots, m) \quad (4.1)$$

where $\underline{D}_{ij}$ represents the $\underline{i}$ th consumers demand for the $\underline{j}$ th commodity, $(p_1, p_2, \dots, p_m)$ the prices of the $\underline{m}$ commodities in the market, and y_i and $\underline{i}$ th consumer's income.

If we are concerned solely with the market for a particular commodity, say $\underline{Q}_j$, then we are interested in the behavior of the demand for $\underline{Q}_j$ with respect to a change in its prices p_j . From the previous chapter it is clear that in order to determine the rate of change of one variable with respect to a shift in a single parameter value one can only do so if all other variables and parameters are treated as constants. Thus, to determine the rate of change of the $\underline{i}$ th consumer's demand for $\underline{Q}_j$ with respect to a shift in p_j all other prices and quantities must be treated as constants. Consequently, the demand for $\underline{Q}_j$, although still depending upon the prices of the other commodities and the consumer's income, becomes a function of p_j alone, i.e.:

$$\underline{D}_{ij} = \underline{D}_{ij}(p_j) \quad .$$

Since the $\underline{i}$ th consumer within a competitive market is indistinguishable from any other consumer with respect to a seller the demand for $\underline{Q}_j$ for all consumers is solely a function of the price of this commodity. Further, since there are $\underline{n}$ consumers each of which demands a certain amount of $\underline{Q}_j$ at a

specific price p_j , the total market demand for Q_j is given by the sum of the individual quantities demanded:

$$D_j = \sum_{i=1}^n D_{ij}(p_j) = D_j(p_j)$$

or by dropping the product subscript

$$D = \sum_{i=1}^n D_i(p) = D(p) \quad (4.2)$$

where (2) is derived by holding all other prices and the incomes of all n consumers constant.

For example, consider a market which only contains two commodities Q_1 and Q_2 ^{1/}. If the i th consumer spends all his income on these two commodities then his expenditure can be represented by:

$$y_i = p_1 q_1 + p_2 q_2 \quad (4.3)$$

where p_1 and p_2 are the prices and q_1 and q_2 are the amounts purchased of Q_1 and Q_2 respectively. Under equilibrium conditions the quantity demanded by the i th consumer for each item is directly deduced as:

$$D_{i1} = q_1 = \frac{y_i^0}{C_1 p_1} \quad \text{and} \quad D_{i2} = q_2 = \frac{y_i^0}{C_2 p_2} \quad (4.4)$$

^{1/} While a two commodity market may appear somewhat unrealistic to the reader, realism can be introduced by considering one of these items as a composite commodity defined in the Hicksian sense. Under this definition Q_2 , for example, includes all other commodities including savings so that the total amount spent on Q_1 and Q_2 represents the consumer's total income for the period. For further discussion of this point see: J.R. Hicks, A Revision of Demand Theory, Clarendon Press, Oxford, 1956; and G.P.E. Clarkson, op. cit., Chapter 3.

where C_1 and C_2 are constants and where D_{11} and D_{12} can be seen to be monotonically decreasing functions of price.^{2/} By summing over all consumers in the market the result is a market demand curve for each commodity, i.e.:

$$D_1 = Q_1 = \frac{Y^0}{C_1 P_1} \qquad D_2 = Q_2 = \frac{Y^0}{C_2 P_2} \qquad (4.5)$$

where Y^0 is the total equilibrium level of income for the n consumers. Since (4.4) is derived by holding income and the other price constant, the relations in (4.4) are clearly a function of the single price, p_1 or p_2 . Similarly the aggregate demand functions of (4.5) are also monotonically decreasing functions of one price alone.

The second main hypothesis of the market theory is that the function which represents the amount a firm will produce of a particular item depends solely upon the market price for this item. Under the equilibrium postulates a firm produces at the point where marginal cost equals market price. Thus, the supply function of the firm is identical to a portion of the function representing the marginal cost curve of the firm for that item. Under conditions described as the "short-run", where each firm may vary output but is unable to vary the size of its plant, the relevant segment is that portion of the marginal cost curve which lies above the firm's average variable cost curve.

^{2/}To demonstrate that demand curves are monotonically decreasing functions of price under certain general conditions it is necessary to take into account the Slutsky relation incorporating the income and substitution effects. This relation is derived in: E.E. Slutsky, "On the Theory of the Budget of the Consumer," Giornale degli Economisti, Vol. 51, July, 1951, pp. 1-26. (Reprinted in, American Economic Association, Readings in Price Theory, Irwin, Homewood, Illinois, 1952, pp. 27-56. For an excellent presentation of the proof of the monotonicity of demand curves see: J.M. Henderson and R.E. Quandt, Microeconomic Theory, McGraw-Hill, New York, 1958, Chapter 2.

Since the marginal cost curve measures the quantity a firm will supply of a particular item over a range of market prices, the short-run marginal cost curve is a function of output and price. Hence, the relevant portion of the i th firm's marginal cost curve for a single item can be represented by:

$$MC_i = \phi_i(q_i) \quad (4.6)$$

where MC_i is the first derivative of the i th firm's total cost of producing the item taken with respect to the output.^{3/} Under the short-run condition the supply function for the i th firm is obtained by applying the first order condition for equilibrium and by solving for q_i . From the first order condition the relation $p=MC_i$ is inferred. And by solving for q_i in terms of a supply function S_i we get:

$$\begin{aligned} S_i &= S_i(p) && \text{for all } p \text{ that lie above the minimum point on the} \\ &&& \text{average variable cost curve} \\ S_i &= 0 && \text{for all } p \text{ that are below the minimum point on the} \\ &&& \text{average variable cost curve.} \end{aligned} \quad (4.7)$$

The aggregate supply function for commodity Q is obtained in the same manner as the aggregate demand function. All prices and costs of other commodities are held constant so that the sum of the n individual supply function represents the aggregate:

$$S = \sum_{i=1}^n S_i(p) = S(p) \quad (4.8)$$

By applying the second order conditions for equilibrium it can be shown that in

^{3/} If the i th firm cost function for item q_i is represented by $C=\phi(q_i)+b$, i.e. a function of the level of output plus the cost of fixed inputs, $MC_i=\frac{dC}{dq_i}=\phi'(q)$ since b is a constant.

most normal^{4/} circumstances the rate of change of output with respect to changes in price is positive. Further that the ith firm's supply function is a monotonically increasing function with respect to price. Thus, due to the process by which it is derived, the aggregate supply function is also a monotonically increasing function.

Under conditions described as the "long-run" the firm is regarded as being able to adjust the values of all variables. In the long-run, then, the supply function is identical with that portion of the marginal cost curve which lies above the firm's average cost curve. The long-run supply function is derived under the same conditions as above and for the ith firm it can be represented by:

$$S_i = S_i(p) \quad (i=1,2,\dots,n)$$

Consequently, the long-run, aggregate supply function is given by:

$$S = \sum_{i=1}^n S_i(p) = S(p) \quad (i=1,2,\dots,n) \quad (4.9)$$

2. External Economies and the Slope of the Supply Function

In the preceding analysis it is noted that in most normal circumstances the supply function can be shown to have a positive slope. As part of the hypothesis about the supply function the firm's costs are assumed to be solely a function of its output. Earlier the ith firm's cost function for a specific item was represented by:

^{4/}See: J.M. Henderson and R.E. Quandt, op. cit., p. 90.

$$C = \Phi(q) + b \quad (4.10)$$

where b represents the fixed costs and $\Phi(q)$ the variable costs associated with a particular level of output. In this case, the function $\Phi(q)$ reflects the labor, material, and other costs which vary with the output level. If all these items do not depend upon the level of output or behavior of any other firm, then the supply function for the i th firm can be shown to have a positive slope.

However, it frequently happens that a firm's total costs do depend upon the output of other firms. Other firms may develop new methods of production which allow them to increase output as well as reduce price. If the i th firm employs these items within its manufacturing process its cost function will reflect the lower input costs. Similarly if new industries arise which use many of the same raw materials as firm i , the enlarged demand for these inputs may raise their market price. Accordingly, the increased costs will be reflected in the i th firm's cost function. As a result there are two possible cases where the cost function of firm i is affected by the output levels of other firms. The first is where an expansion of, say, the j th firm's output lowers the total cost function of the i th firm--a situation where external economies are realized. The second distinct possibility is when the expansion of the j th firm's output raises the total cost function of the i th firm--a situation where external diseconomies are realized.^{5/}

^{5/}A clear discussion of the classical problem posed by external economies and diseconomies is presented in: J.E. Meade, "External Economies and Diseconomies in a Competitive Situation," Economic Journal, Vol. LXII, March, 1952, pp. 54-67. For an excellent analysis of the various welfare solutions that are possible under these circumstances see: O.A. Davis and A. Whinston, "Externalities, Welfare, and the Theory of Games," Journal of Political Economy, Vol. LXX, June, 1962, pp. 241-262.

In order to inspect the effect external economies and diseconomies have upon the slope of the supply function it is necessary to represent the i th firm's cost function as dependent upon the output levels of all n firms:

$$C_i = \Phi_i(q_1, q_2, \dots, q_n) \quad (i=1, 2, \dots, n)$$

where q_i is the output of the i th firm. At the same time the profit of the i th firm is the difference between the total revenue derived from the sale of commodity q_i and the cost of producing this output C_i . If the market price for q_i is given by p , then the i th firm's profit is

$$\Pi_i = p q_i - C_i \quad (i=1, 2, \dots, n) \quad (4.11)$$

For the n firms in this commodity market there are n functions represented by (4.11). In order to discover the equilibrium solution we proceed as outlined in the previous chapter. First we apply the first order condition which in this case requires that we differentiate Π_i with respect to q_i and set this partial differential equal to zero. In order to take the partial differential of Π_i with respect to q_i it must not be forgotten that at the same time the values of all other variables are held constant. Proceeding with this application of partial differentials to the system of functions in (4.11) we have:

$$\begin{aligned} \frac{\partial \Pi_1}{\partial q_1} &= p - \frac{\partial C_1}{\partial q_1} = p - \frac{\partial \Phi}{\partial q_1} (q_1, q_2, \dots, q_n) = 0 \\ \frac{\partial \Pi_2}{\partial q_2} &= p - \frac{\partial C_2}{\partial q_2} = p - \frac{\partial \Phi}{\partial q_2} (q_1, q_2, \dots, q_n) = 0 \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ \frac{\partial \Pi_n}{\partial q_n} &= p - \frac{\partial C_n}{\partial q_n} = p - \frac{\partial \Phi}{\partial q_n} (q_1, q_2, \dots, q_n) = 0 \end{aligned} \quad (4.12)$$

By imposing the second order conditions that

$$\frac{\partial^2 \phi_i}{\partial q_i^2} (q_1, q_2, \dots, q_n) > 0 \text{ for all } (i=1, 2, \dots, n)$$

upon the system of relations in (4.12) the unique, equilibrium solution to this set of functions is generated in terms of q_i . By writing these solutions in terms of $\underline{S}_i$ instead of q_i the following system of solutions results:

$$\begin{aligned} S_1 &= S_1(p) \\ S_2 &= S_2(p) \\ &\vdots \\ S_n &= S_n(p) \end{aligned} \tag{4.13}$$

Now, if the slopes of $\underline{S}_1, \underline{S}_2, \dots, \underline{S}_n$ are known then it is possible to infer the slope of the aggregate supply function

$$S = \sum_{i=1}^n S_i(p) = S(p) \quad .$$

But each firm bases its output decision upon the relevant portion of its own marginal cost function. Hence, each firm, for example the i th firm, observes the equilibrium output values of all the other firms $(q_1^0, q_2^0, \dots, q_h^0, q_j^0, \dots, q_n^0)$ and selects that value of q_i for his own output which satisfies his own equilibrium relation:

$$\frac{\partial \pi_i}{\partial q_i} = p - \frac{\partial \phi_i}{\partial q_i} (q_1^0, q_2^0, \dots, q_h^0, q_i, q_j^0, \dots, q_n^0) = 0 \quad .$$

Similarly, the equilibrium value of q_i^0 may affect the output decisions of other entrepreneurs. Hence, while the supply functions (4.13) state that, after all these adjustments are made, the equilibrium outputs are a function of price they do not include a reference as to the sign of their respective slopes.

To determine the slopes of the respective supply functions one either needs to know the signs of all the coefficients in (4.12) or one must apply the methods

of comparative statics. In the latter case one would apply the schema outlined in the previous chapter and by a suitable application of restrictions infer the directions of the rates of change of the various quantities to their respective prices. Of course, this analysis can only be carried out if the equilibrium solutions to the system of relations in (4.12) are known. But presuming that these relations have been solved then this solution can be employed as the initial conditions in order to proceed step by step to the resolution of the signs of the individual rates of change. If each of the supply functions $\underline{S}_i$ has a slope with the same sign, then the aggregate function $\underline{S}$ will also have the same sign. But if not all $\underline{S}_i$ are either positive or negative then without knowing the relative magnitude of each component we are still unable to determine the sign of $\underline{S}$.

In the former case, however, we know the signs of all the coefficients in (4.12). If we represent the cost functions of the n firms in the following way

$$\begin{aligned} C_1 &= a_{11}q_1^2 + a_{12}q_2^2 + \dots + a_{1n}q_n^2 \\ C_2 &= a_{21}q_1^2 + a_{22}q_2^2 + \dots + a_{2n}q_n^2 \\ &\vdots = \quad \quad \quad \vdots \quad \quad \quad \vdots \\ C_n &= a_{n1}q_1^2 + a_{n2}q_2^2 + \dots + a_{nn}q_n^2 \end{aligned}$$

then the coefficients, whose signs we know, are represented by $\underline{a}_{11}, \underline{a}_{12}, \underline{a}_{13}, \dots, \underline{a}_{nn}$. If the market can be characterized entirely in terms of external economies--i.e., an expansion in the output of each firm lowers the total cost function of at least one other firm--then all $\underline{a}_{ij}$ for $(i \neq j)$ must be negative. At the same time, of course, we need to know the signs of the remaining $\underline{a}_{ij}$ when $(i=j)$. If they

are all of one sign, say positive, then we can proceed to a unique solution in the following way. From (4.12) the equilibrium relations are given by:

$$\frac{\partial \pi_1}{\partial q_1} = p - 2a_{11}q_1 = 0$$

$$\frac{\partial \pi_2}{\partial q_2} = p - 2a_{22}q_2 = 0$$

$$\vdots \quad \vdots \quad \vdots$$

$$\frac{\partial \pi_n}{\partial q_n} = p - 2a_{nn}q_n = 0$$

Solving these equations into the form given in (4.13) one gets:

$$s_1 = \frac{p}{2a_{11}}$$

$$s_2 = \frac{p}{2a_{22}}$$

$$\vdots \quad \vdots$$

$$s_n = \frac{p}{2a_{nn}}$$

where the aggregate supply function is

$$S = \sum_{i=1}^n s_i = \frac{p}{2} \left(\frac{1}{a_{11}} + \frac{1}{a_{22}} + \dots + \frac{1}{a_{nn}} \right) \quad (4.14)$$

If, as supposed, all $\underline{a}_{ij}$, ($i=j$), are positive then it is clear that the slope of (4.14) is also positive. Similarly, if all $\underline{a}_{ij}$, ($i=j$), are negative then the slope of (4.14) is negative. But, if not all of the $\underline{a}_{ij}$, ($i=1$), are either positive or negative then once again we need to know the relative magnitude of each $\underline{a}_{ij}$, ($i=j$), before the slope of the resulting aggregate supply function can be determined.

3. Market Equilibrium

The equilibrium of a commodity market is realized when the quantity demanded through the aggregate demand function is equal to the quantity supplied through the aggregate supply function. Since both functions are stated in terms of the market price for a particular commodity, one condition for equilibrium is realized if one price prevails for that commodity throughout the market. Consequently, for market equilibrium of a particular commodity,

$$D(p) = S(p) \tag{4.15}$$

Under short-run conditions sellers can vary output but not their plant and equipment. Hence, if under these conditions (4.15) does not hold for some price $p = p_1$, then either buyers wish to purchase more of the commodity than is available at p_1 or sellers wish to sell more than is being purchased at this price. If such an event occurs then various efforts can be expected to be made to alter the price p_1 so that both buyers and sellers are making consistent demands upon each other.

For example, suppose the production facilities of each supplier are such that the requisite quantity of the commodity in question can be produced in a very short period of time. Further, suppose that buyers and sellers enter the market for the purpose of making contracts to buy and sell certain quantities at specific prices. If these contracts are such that they entitle both parties to recontract if either or both parties are able to find more favorable offers, then the process of contracting and recontracting will not cease until each buyer and supplier is completely satisfied. If an auctioneer records the price and quantity of each contract as it is made so that all participants are aware of all bids and offers, then because of the equilibrium conditions imposed upon both buyers and sellers the contracting process will

proceed in one of two possible ways.

If the price of the initial contract, p_1 , is such that buyers are unwilling to purchase at this price all the items that suppliers indicate they are willing to supply, some suppliers will attempt to recontract at a lower price, p_2 . Once p_2 is announced by the auctioneer all contracts at p_1 are renegotiated and suppliers will soon see whether they are able to sell all that they are willing to supply at the new price. If they are unable to do so recontracting will take place at a yet lower price, p_3 . This process continues until such time when the contract price is just sufficient to allow the suppliers to sell all the items they are willing to produce at this price. Such a price is, of course, the equilibrium price, p^0 . Once all contracts are made at this price, suppliers produce the required output and both consumers and sellers are satisfied.

If, however, the initial contract price, p_1 , is such that consumers are willing to contract for more items than the suppliers are willing to produce at this price, then the recontracting process works in the opposite direction. Now, it is some of the consumers who offer a higher price, p_2 , which then becomes the price at which new contracts are made. At p_2 suppliers are willing to increase the amount they are willing to produce. If this amount is not sufficient to satisfy the consumers demands recontracting continues until the auctioneer announces a price, p^0 , at which both consumers and suppliers are satisfied.

Despite the restrictive nature of this example it is clear that the equilibrium conditions imply that the market for a particular commodity, in the short-run, is in equilibrium if and only if there is one price which simultaneously satisfies both consumers and producers. Consequently, within any

particular market the equilibrium price is determined by solving the demand and supply functions (4.2) and (4.8) under the constraint $D(p) = S(p)$.

For long-run equilibrium each producer may alter his output as well as his plant and equipment. At the same time his supply function is now equivalent to that portion of his marginal cost curve which is above his long-run, average cost curve. Consequently, for a market in which there are a specific number of firms the equilibrium price for this market will be given by the intersection of the aggregate, long-run supply and demand functions, i.e., $D(p) = S(p)$.

In this case, however, we must take into account the amount of profit each firm is making at the equilibrium price. Since we started with a specific number of firms the equilibrium price may be such as to allow some or all of these firms to receive more than the minimum returns necessary to them to remain in business. If such is the situation then new firms will be enticed to enter this market. At some market price, p_1 , the profits accruing to an efficient firm will be such that new firms will enter this market. Once this occurs the aggregate supply function will be increased. As a result, the market price, p_2 , necessary to clear this larger amount will be lower than p_1 . As long as new producers enter the market the recontracting process will permit the price to fall until a point is reached where excess or above normal profits can no longer be earned. At this point long-run demand equals long-run supply and excess profits are zero. Hence, for long-run equilibrium in a single commodity market there are two conditions which must be satisfied:

$$D(p) = S(p)$$

and

$$\Pi = p \sum_{i=1}^n q_i - \sum_{i=1}^n C_i = 0 \quad (4.16)$$

Substituting $\sum q_i = S$ and $\phi_i(q_i)$ for C_i (4.16) becomes

$$\Pi = pS - \sum_{i=1}^n \phi_i(q_i) = 0$$

But if there are n firms in the industry each producing q_i , the total cost for the i th firm of producing q_i , if all firms have identical cost functions, can be represented by $\phi_i(\frac{S}{n})$. Substituting $\phi_i(\frac{S}{n})$ for $\phi_i(q_i)$ we get:

$$\Pi = pS - \sum \phi_i(\frac{S}{n}) = 0$$

or

$$\Pi = pS - n\phi_i(\frac{S}{n}) = 0 \quad (4.17)$$

Consequently, by applying the equilibrium conditions (4.15) and (4.17) to the long-run supply and demand functions we can not only determine the equilibrium price for a particular commodity, but also the number of firms in the industry, as long as all these firms have identical cost functions.

A. An Example^{6/}

In the preceding discussion it is implicitly supposed that consumers and producers are sufficiently near to each other so that no reference is made to transportation or other types of marketing costs. In many markets, however, consumers and producers are not spatially near one another and some producers have to ship their products greater distances than others. If such a market is to be in equilibrium only one price may prevail. As a result, since some firms have greater transportation costs than others, this difference in costs must affect

^{6/}This example is taken from J.M. Henderson and R.E. Quandt, op. cit., pp. 101-104.

the amount they are willing to supply. To explore the manner in which firms within such a market behave--city markets for dairy products are representative examples--we apply the equilibrium analysis in the following way.

First we need to include the transportation costs in the total cost function of the representative firm:

$$C_i = \phi_i(q_i) + b_i + \alpha_i q_i \quad (4.18)$$

where $\phi_i(q_i)$ and b_i represent, as before, the variable and fixed cost components, and where α_i is the cost per unit of q_i for transporting it from the firm to the city market. The profit for the i th firm is given by:

$$\Pi = pq_i - \phi_i(q_i) - b_i - \alpha_i q_i \quad (4.19)$$

By applying the first order conditions for equilibrium to (4.19) we have:

$$\frac{\partial \Pi_i}{\partial q_i} = p - \frac{\partial}{\partial q_i} \phi_i(q_i) - \alpha_i = 0$$

or

$$\frac{\partial}{\partial q_i} \phi_i(q_i) = p - \alpha_i \quad (4.20)$$

Accordingly, to be operating at equilibrium output (4.20) states that the firm should equate the marginal cost of production to the market price minus the transportation cost. If each of the n firms in the industry has the same production cost function $\phi_i(q_i)$ but has a different transportation cost α_i , then the marginal cost of production will differ for each firm. Thus, for a specific equilibrium price, p^0 , each firm will be willing to supply a different amount. In the short-run, the market will be in equilibrium when the total supply equals the total demand.

To make the example more specific let us suppose that the firms supplying commodity Q to a market can be divided into two classes: those whose

transportation costs are 10 dollars a unit and those whose costs are 13 dollars a unit. There are one hundred firms which are evenly divided between the two classes. If all firms have the same production cost functions, $\phi_i(q_i)$, then the total costs for the representative firms, where $\phi_i(q_i) = 0.5q_i^2 + \alpha_i q_i$, are given by:

$$C_1 = 0.5q_1^2 + 10q_1 \qquad C_2 = .05q_2^2 + 13 q_1$$

where the subscripts 1 and 2 represent firms of the two classes. Applying the first order conditions for equilibrium we get from (4.19) and (4.20):

$$q_1 = p - 10 \qquad q_2 = p - 13$$

The supply functions for these firms are derived in the same manner as (4.7) by substituting $\underline{s}_i$ and $\underline{q}_i$, and remembering that $\underline{s}_i = 0$ for all points on the marginal cost curve which lie below the minimum point on the average variable cost curve. Under these conditions the representative supply functions are:

$$\begin{array}{lll} S_1 = 0 & \text{if} & 0 \leq p < 10 \\ S_1 = p - 10 & \text{if} & 10 \leq p \\ S_2 = 0 & \text{if} & 0 \leq p < 13 \\ S_2 = p - 13 & \text{if} & 13 \leq p \end{array} \qquad (4.21)$$

Hence, firms in the first class will not supply any output if the price is less than 10 dollars, while the second class of firms will not supply any output if the price is less than 13 dollars a unit.

The aggregate supply function is derived as in (4.8). Since there are one hundred firms this function is given by the following relations:

$$\begin{aligned} S &= 0 & \text{if} & \quad 0 \leq p < 10 \\ S &= 50(p-10) = 50p - 500 & \text{if} & \quad 10 \leq p < 13 \\ S &= 50(p-10) + 50(p-13) \\ &= 100p - 1150 & \text{if} & \quad 13 \leq p \end{aligned}$$

To solve for the equilibrium price level we need to know the aggregate demand function. Suppose this function can be represented by

$$D = 30p + 250 .$$

Then the equilibrium price is determined by setting $\underline{S} = \underline{D}$, or:

$$100p - 1150 = 30p + 250$$

$$p = 20$$

and

$$S = D = 850$$

From (4.21) each firm in class 1 sells 10 units while each firm in class 2 sells 7. From the profit relations,

$$\Pi_1 = pq_1 - C_1(q_1) \qquad \Pi_2 = pq_2 - C_2(q_2)$$

we can determine that each firm in class 1 makes a profit of 50 dollars while each firm in class 2 makes a profit of 20 dollars. In the short-run this result constitutes an equilibrium position.

For a long-run equilibrium, however, we need to impose the additional constraint that profits must be equal to zero. In this particular example long-run equilibrium could be achieved in a variety of ways. One way would be for new firms to enter the geographical location of the class 1 firms until such time as the price per unit falls below 13 dollars. At this point the firms in class 2 would no longer produce and the market would be composed solely of class 1 firms. Another possibility would be for the owners of the land

nearer the market to charge a higher rent than that being charged the class 2 firms. As new firms enter the market, to take advantage of the higher profits of the class 1 location, rents would rise until such time as they made the net profit from the two locations the same. While neither of these processes need necessarily lead the market to a point of zero profits, some combination of these and other similar processes must take place if the market is to reach a long-run equilibrium position.

4. Static and Dynamic Stability of Equilibrium

The previous section is concerned with the process by which equilibrium, both short and long-run, is attained. In describing the process, however, it was noted that the opening price on any market need not be the equilibrium price. Further, once equilibrium is reached a shift in consumer preferences or a shift in the supply curve, caused by some technological or other change, can alter the equilibrium. Since there are a variety of factors which can disturb the equilibrating process two possibilities are evident: the first is that the disturbing factors may prevent an equilibrium from ever being attained; the second is that once it is arrived at the equilibrium point may not be a stable one. If an equilibrium is stable then the market will return to equilibrium no matter what disturbance has affected it. But, if an equilibrium is unstable then the market will not return to equilibrium once it has reached and been disturbed from this equilibrium point. Consequently, it is necessary to inquire into the properties of equilibrium points, both in static and dynamic cases, so as to be able to distinguish those that are stable from those that are not.

A. The Static Conditions

In the last section it is shown that a market equilibrium for a particular commodity can be achieved if after each set of contracts are made both buyers and suppliers can recontract if and when more favorable opportunities become available. If the initial price is such that consumers are willing to contract for more of the commodity than the sellers are willing to supply at that price, the price is increased by some consumers to enable them to increase their purchases. Under a static analysis there is no concern over the path described by this process over time. Instead we merely wish to know the direction of each change or adjustment, and whether this is toward or away from the equilibrium point.

In order to be able to examine the adjustment process it is necessary to introduce the notion of excess demand at a particular price. If the price prevailing at any instant in a market is not equal to the equilibrium price then there is excess demand at that price. If the prevailing price is above (below) the equilibrium point the excess demand is negative (positive). In figure 1 there is a negative excess demand at price p_2 , and a positive excess demand at price p_1 . Due to the condition for market equilibrium, e.g., $D(p) = S(p)$, there clearly cannot be any excess demand at price p^0 . Consequently the excess demand at any price p can be represented by:

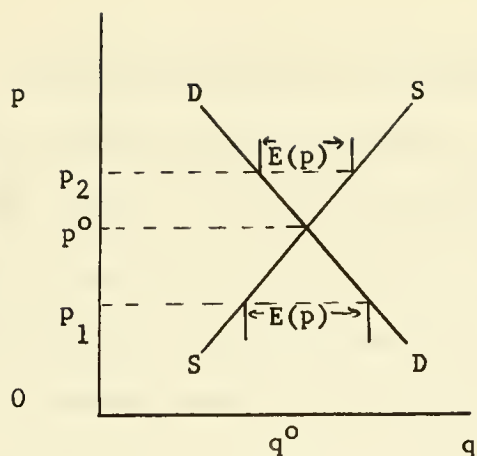


Figure 1

$$E(p) = D(p) - S(p) \quad (4.22)$$

Suppose, for the moment, that there is excess demand in the market under consideration, then the question to be answered concerns the direction of the rate of change of $\underline{E}(p)$ with respect to changes in p . If the price is at p_1 and shifts toward p^0 then, because $\underline{E}(p)$ decreases with

this shift in p , $\frac{dE(p)}{dp} < 0$. If the price shifts from p_2 towards p^0 then, for the same reason, $\frac{dE(p)}{dp} < 0$. Hence, as long as the rate of change of $\underline{E}(p)$ with respect to change in the market price is negative the market is moving toward an equilibrium.

To ensure that this condition is met an additional constraint, or postulate, is imposed upon the behavior of the consumers in the market. This constraint, which is known as the Walrasian stability condition, requires buyers to raise their bids if excess demand is positive and sellers to lower their prices if it is negative. If buyers and sellers behave in such a manner so as to satisfy this condition then as long as a shift in price reduces the excess demand the market is stable. The Walrasian stability condition is given by:

$$\frac{dE(p)}{dp} = \frac{d}{dp} D(p) - \frac{d}{dp} S(p) < 0 \quad (4.23)$$

So far the stabilizing process has been regarded as one which is solely a function of shifts in the market price. However, in the analysis of market equilibrium suppliers may respond to shifts in prices by raising or lowering the amount they are willing to supply. As a result, we must inquire into the conditions under which these shifts in the amount supplied affect the stability of the market.

To isolate the effects of the shifts on the supply side the notion of an excess demand price is introduced. If at any point of time in a particular market there is a difference between the price consumers are willing to pay and the price sellers are asking for a specific quantity of a commodity then this difference is the excess demand price.

An excess demand price can occur in the following way. If we take the aggregate demand and supply functions, $\underline{D}(p)$ and $\underline{S}(p)$, and restate them in terms of the quantity (not price) required for market equilibrium, the equilibrium condition is $\underline{D}=\underline{S}=q$. At equilibrium there is one price, p^0 , at which consumers will purchase this equilibrium quantity q^0 . But, at any point other than at q^0 , there is a price, p_s , at which suppliers are willing to sell this quantity. At the same time there is also a price, p_d , at which consumers are willing to buy such a quantity. At all points other than at equilibrium the supply price, p_s , is not equal to the demand price, p_d . To determine these prices for a specific market one solves the aggregate demand and supply functions in terms of p_s and p_d as follows:

$$p_d = D^{-1}(q)$$

$$p_s = S^{-1}(q)$$

where D^{-1} and S^{-1} are the inverses of the aggregate demand and supply functions.^{7/} Once the demand and supply prices are known, the excess demand price is given by the difference between the two, or:

$$E(q) = p_d - p_s = D^{-1}(q) - S^{-1}(q) \quad (4.24)$$

^{7/}If for the relation $y=f(x)$ there is a solution it can be written as $x=f^{-1}(y)$, where f^{-1} is what is called the inverse of $f(x)$. For further discussion see Appendix A.

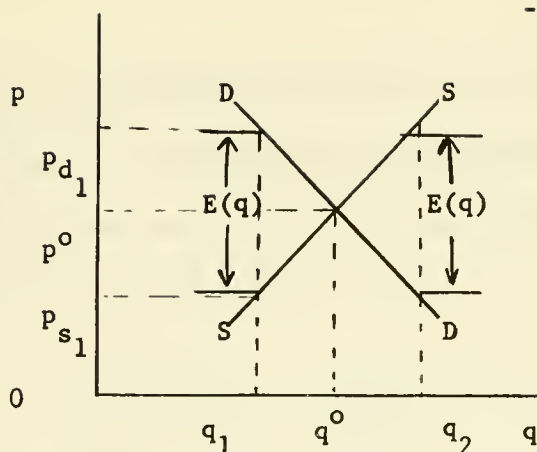


Figure 2

The stability of the market depends, then on the direction of the rate of change of $\underline{E}(q)$ with respect to a shift in q . If for a certain market the quantity to be supplied is represented by q_1 , in Figure 2, the demand and supply prices are given by

p_{d1} and p_{s1} . If the quantity supplied

shifts from q_1 toward q^0 then we are interested in the directional shift $\underline{E}(q)$.

Since in this instance, $\underline{E}(q)$ will decrease as q_1 shifts toward q^0 , $\frac{dE(q)}{dq} < 0$.

Similarly, if the quantity supplied is represented by q_2 and then shifts toward

q^0 , $\frac{dE(q)}{dq} < 0$. Accordingly, the market will approach a stable equilibrium as

long as the rate of change of the excess demand price is negative with respect

to changes in the quantity supplied. To ensure that this condition is satisfied

producers must be willing to increase output if $\underline{E}(q) > 0$, e.g., such as when

q_1 represents the amount they are currently supplying. At the same time they

must also be willing to decrease output if $\underline{E}(q) < 0$ --e.g., when at a position

represented by q_2 . This requirement on the behavior of producers is known as

the Marshallian stability condition. Consequently, the market is in stable

equilibrium in the Marshallian sense if

$$\frac{dE(q)}{dq} = \frac{d}{dq} D^{-1}(q) - \frac{d}{dq} S^{-1}(q) < 0 \quad (4.25)$$

If the aggregate demand function has a negative slope and the aggregate supply function has a positive slope both conditions, (4.23) and (4.25), are satisfied. Manifestly, under these conditions the market is stable from both points of view. However, if external economics or diseconomics are present

within the market the supply function may have a negative slope. Under this eventuality both equilibrium conditions cannot be simultaneously satisfied.

To show that such is the case divide both sides of (4.25) by

$[\frac{d}{dq} D^{-1}(q)] \cdot [\frac{d}{dq} S^{-1}(q)]$ to get

$$\frac{1}{\frac{d}{dq} S^{-1}(q)} - \frac{1}{\frac{d}{dq} D^{-1}(q)} < 0 \quad (4.26)$$

But

$$\frac{1}{\frac{d}{dq} S^{-1}(q)} = \frac{d}{dp} S(p) \quad \frac{1}{\frac{d}{dq} D^{-1}(q)} = \frac{d}{dp} D(p)$$

By substituting these values into (4.26) we get:

$$\frac{d}{dp} S(p) - \frac{d}{dp} D(p) < 0 \quad (4.27)$$

But the Walrasian stability condition (4.23) states that:

$$\frac{d}{dp} D(p) - \frac{d}{dp} S(p) < 0$$

Clearly both cannot be satisfied at one time. Thus, if the equilibrium is stable in the Walrasian sense it is unstable in the Marshallian, and vice versa.

For example, consider the situation represented in Figure 3. At price, p^0 ,

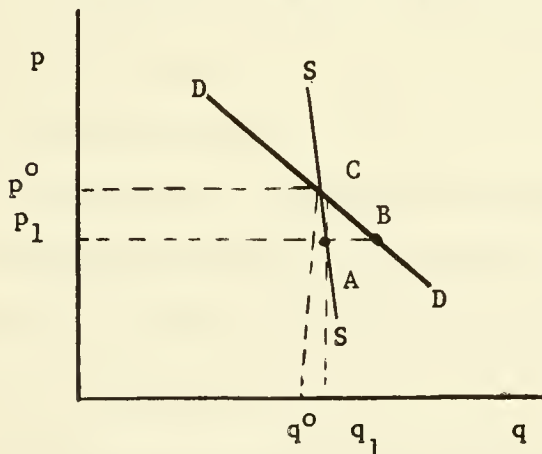


Figure 3

the amount demanded equals the amount supplied, q^0 , and the market is in equilibrium. But, if a disturbance occurs and the price temporarily shifts to p_1 there will be a positive excess demanded represented by AB. Under the Walrasian conditions buyers will tend to raise their prices and the excess demand will be reduced.

Concurrently, however, suppliers are willing to produce q_1 at p_1 which implies a positive, excess demand price represented by AC. As a result, under the Marshallian conditions, producers will tend to increase the supply. If the supply increases consumers will be unsuccessful in their attempts to reduce the excess demand by raising the price, and the actual price and quantity will move away from the equilibrium point.

B. The Dynamic Conditions

In the dynamic case we are interested in examining the time path of the process by which equilibrium is reached. If we revert to the example where consumers contract with suppliers at one price and then recontract at a new price the minute a more favorable contract can be made, we are now interested in path these successive prices describe over time. This recontracting process is dynamically stable if, over time, the price approaches the equilibrium price. It is dynamically unstable if the direction in which the prices change is away from the equilibrium price. Stated in this manner dynamic stability has been defined in a Walrasian sense. Clearly, it is also possible to examine the path described by the recontracted quantities. If these approach the equilibrium quantity over time then the market will be dynamically stable in the Marshallian sense.

To explore the conditions under which an equilibrium is dynamically stable the Walrasian position is adopted here. Accordingly, the analysis is based upon a study of the effects of excess demand upon price.^{8/} As has already been

^{8/} The same analysis can be applied under Marshallian conditions but is omitted here for brevity.

pointed out consumers respond to an excess demand by raising the price. If each price change is considered to take place within a discrete interval of time, this process can be represented by:

$$a E(p_{t-1}) = p_t - p_{t-1} \quad (4.28)$$

where a is a positive constant, p_t the price at period t , and p_{t-1} the price at the period before t . While there are numerous different relations which will express a similar type of behavior, (4.28) states that a positive excess demand in period $t-1$ will induce a price increase in period t . Suppose for the moment that the aggregate demand and supply functions are known and that they can be represented by the relations:

$$D_t = b p_t + c \quad (4.29)$$

$$S_t = d p_t + e \quad (4.30)$$

Then from relation (4.22) excess demand in period $t-1$ is:

$$E(t-1) = (b p_{t-1} + c) - (d p_{t-1} + e) = (b-d)p_{t-1} + c - e \quad (4.31)$$

By substituting this result into (4.28) we get:

$$a[(b-d)p_{t-1} + c - e] = p_t - p_{t-1} \quad (4.32)$$

$$[1 + a(b-d)]p_{t-1} + a(c-e) = p_t$$

which represents the behavior of the price, p , as it shifts from one period to the next.

At equilibrium the aggregate demand equals the aggregate supply, and excess demand is equal to zero. Therefore, by setting $D_t = S_t$ in (4.29) and (4.30) we can solve for the equilibrium price $p_t = p^0$ as follows:

$$p^0 = \frac{c-e}{d-b} \quad (4.33)$$

From (4.32) we know the path that the market prices will describe (as long as the relation (4.31) is satisfied by the market). And from (4.33) we

know the value of the equilibrium price, p^0 , in terms of the constants in the aggregate demand and supply functions (4.29) and (4.30). What has yet to be determined is whether and under what conditions p_t will approach p^0 as time increases.

To be able to answer these questions it is necessary to solve the relation (4.32) in terms of p_t . But, to solve this first order difference equation we need to know the initial value of p at $t=0$ ^{9/}. Suppose for the moment that $p=p_1$ at $t=0$, then the solution of (4.32) in terms of p_t is given by:

$$p_t = (p_1 - \frac{c-e}{d-b}) [1 + a(b-d)]^t + \frac{c-e}{d-b} \quad (4.34)$$

Since the last term in (4.34) is the value of the equilibrium price, the equilibrium is dynamically stable if $p_t \rightarrow \frac{c-e}{d-b}$ as t increases. This will occur--i.e. p_t will converge directly to p^0 if

$$0 < 1 + a(b-d) < 1 \quad (4.35)$$

Now, if the aggregate demand function has a negative slope, $b < 0$, and if the aggregate supply function has a positive slope, $d > 0$, then under these two conditions $b < d$ which is sufficient to ensure that, $1 + a(b-d) < 1$. For the left hand side of (4.35) to hold

$$a < \frac{1}{d-b} \quad (4.36)$$

where a measures in these relations the degree to which buyers and sellers adjust their bids in the presence of excess demand.

If the aggregate supply function has a negative slope the equilibrium will be dynamically stable as long as $b < d$. For since $\frac{1}{b}$ is the slope of the demand function (4.29) and $\frac{1}{d}$ is the slope of the supply function (4.30), the equilibrium

^{9/}See math Appendix for solution procedure.

is dynamically stable as long as $\frac{1}{b} > \frac{1}{d}$.^{10/}

Oscillations around the equilibrium level are introduced if condition (4.35) is not satisfied and, $1 + a(b-d) < 0$. This can occur if both $\underline{a} < 0$ and $b-d < 0$, and $\underline{a}$ is sufficiently large to make $1 + a(b-d) < 0$. If the values of $\underline{a}$, $\underline{b}$, and $\underline{d}$ are such that $-1 < 1 + a(b-d) < 0$ the equilibrium is dynamically stable and the price converges to p^0 with decreasing oscillations. But, if $1 + a(b-d) < -1$ then the oscillations about p^0 will increase with time and the market will be dynamically unstable.

Dynamic stability in a market implies, of course, that the market is stable under static conditions. The converse, however, does not follow. Thus, a market which satisfies the Walrasian or Marshallian conditions for static stability may at the same time be dynamically unstable. Consequently, when analysing the equilibrium position of a particular market great care must be taken to identify both the type of analysis relevant to the investigation as well as the behavioral conditions which are supposed to control the equilibrating process.

5. General Market Equilibrium

So far the analysis has been concerned with the processes by which equilibrium is attained in a single commodity market. In order to focus on these processes the analysis excluded any interactions between the markets for different commodities. As a result, the quantities of other goods bought and

^{10/} It should not be forgotten that this analysis is dealing with the Walrasian stability conditions. Under Marshallian conditions this would be an unstable equilibrium.

sold at their respective prices are treated as parameters when the equilibrium price and quantity within a single market is being determined. If we were faced with the task of finding the equilibrium positions of n different markets, one way to proceed would be to treat each market as a separate entity. Under this arrangement each factor and its price would be a variable in the analysis of its own market, while it would be considered as a parameter in the analysis of the remaining $n-1$ markets. Proceeding in this manner it is then possible to determine n separate equilibrium values of prices and quantities for the n markets.

But to determine the price for Q_i in the analysis of the market for Q_i the analysis supposes that we already know the prices of all the other commodities ($Q_1, Q_2, \dots, Q_h, Q_j, \dots, Q_n$). Hence, whether we actually know these prices or not, the method of analysis requires that we assign them particular values. Since all markets are interrelated, the only condition under which it is possible to derive the correct equilibrium point for market Q_i is when the equilibrium values for the remaining $n-1$ commodity markets are already determined. Otherwise one is in the position of deriving an equilibrium position for market Q_i which depends on the prices of the other commodities at a time when these prices are not the equilibrium values.

Since the demand for each commodity depends on the prices of all other commodities and the total income of all consumers, these variables cannot be treated as parameters if we wish to determine the equilibrium position in all markets. For example, in the analysis of consumer demand some commodities are

identified in terms of the consumer's behavior as substitutes or complements of each other.^{11/} One commodity is called a substitute for another commodity if the quantity demanded of the first commodity increases as the price of the second increases. However, if the quantity demanded of the first commodity falls as the price of the second rises then these commodities are called complements of each other. Similar relations hold between the factors used by producers as inputs to the production process. Consequently, if we are interested in simultaneously determining the equilibrium values for all markets, these relations must be taken into account. One way of ensuring that these conditions are met is to treat the functions representing all products and prices as a complete system whose simultaneous solution provides the consistent set of equilibrium values we are after.

While it will not serve the purpose of this book to delve into the analysis of general market equilibrium in too great detail, a brief analysis of one example will permit an examination of how the conditions of a single market equilibrium are applied to the case where many markets are dealt with at one time. To keep the analysis as simple as possible the example is used of the case of a general exchange economy.^{12/}

In a general exchange economy each consumer enters the market with a certain amount or stock of one or more of the total number of available commodities. Let q_{ij}^1 represent the amount of commodity j held by the i th

^{11/}See, for example, the discussion on substitutes and complements in: G.P.E. Clarkson, op. cit., Chapters 213.

^{12/}For a more detailed discussion of multi-market equilibrium, to which this section is indebted, see: J.M. Henderson and R.E. Quandt, op. cit., Chapter 5.

consumer when he enters the market. Since each consumer is free to buy and sell at the market price, after an exchange has taken place the i th consumer's holding of q_j can be represented by q_{ij}^2 . Accordingly, the i th consumer's excess demand for q_j is represented by E_{ij} where:

$$E_{ij} = q_{ij}^2 - q_{ij}^1 \quad (j=1,2,\dots,m) \quad (4.37)$$

If excess demand is positive it means that his consumption of q_j exceeded his initial holdings q_{ij}^1 . Similarly, if excess demand is negative it implies that the i th consumer did not need all of his q_{ij}^1 and was able to exchange some of his holdings for other commodities. Since the consumer's income is equal to the value of his initial holdings, and since the consumer cannot exceed his income, the value of his purchases and sales must equal his income. By representing the consumer's utility function in terms of the quantities of the commodities he consumes, his utility function can be restated as a function of his excess demands and initial holdings in keeping with (4.37). Concurrently, the consumer's behavior in the market is bounded by equilibrium conditions which require that the net value of his excess demands must be equal to zero. If we are dealing with a market containing m commodities, the i th consumer's utility function can then be solved under the first and second order conditions. The solution yields the function

$$E_{ij} = E_{ij}(p_1, p_2, \dots, p_m) \quad \begin{matrix} (i=1,2,\dots,n) \\ (j=1,2,\dots,m) \end{matrix} \quad (4.38)$$

which states that the i th consumer's excess demands depend upon the prices of all other commodities. As long as his initial holdings of commodity q_j is greater than zero his excess demand for q_j may be positive or negative depending upon the prevailing prices. Nevertheless, under equilibrium conditions his net excess demands for q_j must be equal to zero.

An aggregate excess demand function for a particular commodity is constructed in the usual way by adding together the individual excess demand functions for the $\underline{n}$ consumers. For $\underline{Q}_j$ the aggregate excess demand function would be

$$\underline{E}_j = \sum_{i=1}^{\underline{n}} \underline{E}_{ij}(p_1, p_2, \dots, p_m) \quad (4.39)$$

From (4.39) the partial equilibrium price of $\underline{Q}_j$ can be determined by setting the other $(\underline{m}-1)$ prices equal to a set of fixed values. Having determined the partial equilibrium price $\underline{p}_j^0$ it can be substituted back into the individual excess demand functions (4.38) to find out the purchases and sales each consumer made under these conditions.

The problem, however, is to determine the simultaneous equilibrium of all $\underline{m}$ markets. The equilibrium condition is that excess demand in all markets must equal zero:

$$\underline{E}_j(p_1, p_2, \dots, p_m) = 0 \quad (j=1, 2, \dots, m) \quad (4.40)$$

This condition represents a system of $\underline{m}$ equations in $\underline{m}$ variables. If all $\underline{m}$ equations are independent and consistent then it is possible to determine the absolute values of the $\underline{m}$ variables or prices. But, the system of equations represented by (4.40) has only $(\underline{m}-1)$ independent equations.^{13/} Consequently, it is not possible to solve directly for the equilibrium prices. Instead one solves for the equilibrium set of price or exchange ratios. This solution is derived by taking the price of one of the commodities (usually called the numéraire) and

^{13/} For a proof of this assertion see: J.M. Henderson and R.E. Quandt, op. cit., p. 132.

dividing the prices of the remaining $(m-1)$ items by this price.^{14/} As a result the number of unknown price ratios is reduced to $(m-1)$, a solution for which is provided by the $(m-1)$ independent equations.

For example, if we select the price of the m th commodity, p_m , as the base price or numéraire, the excess demand functions at equilibrium become:

$$E_j \left(\frac{p_1}{p_m}, \frac{p_2}{p_m}, \dots, \frac{p_{m-1}}{p_m}, 1 \right) = 0 \quad (j=1, 2, \dots, m) \quad (4.41)$$

where the variables are the exchange ratios of the m commodities relative to the price of Q_m . The solution of (4.41) provides the equilibrium price or exchange ratios. Accordingly, substituting these values back into the individual excess demand functions the specific purchases and sales of each individual can be determined.

Given the equilibrium position of a particular general market case we might now wish to examine the conditions under which such an equilibrium is statically as well as dynamically stable. The analysis proceeds in a manner analogous to that of the case of a single market. But now it is necessary to consider the effect that disturbances in one commodity market have on the remaining markets. In brief, stability in the static case is achieved if the total rate of change of excess demand with respect to price, i.e. $\frac{dE_j}{dp_j}$ ($j=1, 2, \dots, m$), is negative for all possible combinations of prices. Dynamic stability requires an analysis of this time paths of all the price movements over time. To effect such an analysis one needs to know the exact relations governing the price adjustments over time. If all prices always approach their equilibrium values then, of course, the market is dynamically stable.

^{14/} This manipulation is possible because demand functions are homogeneous of degree zero.

Chapter 5

THE EMPIRICAL CONTENT OF MARKET THEORIES

A testable theory, as noted in Chapter 2, can be considered as an empirically interpreted deductive system. The basis of this system consists of primitive terms and a set of independent and consistent postulates. From this basis the remainder of the system is developed by the application of the formal definitions and the logical rules of deductive inference. Accordingly the theory itself consists of the conjunction of this basis and all the propositions or hypotheses that can be deduced from it. To determine whether a particular theory meets the criteria of a formal deductive system one can examine the structure of the theory independently of any meanings assigned to its component parts. But, in order to assess the empirical content of a theory it is the interpretive rules that are the object of the investigation since it is assumed that the deductive system already satisfies the deductive criteria. In the following analysis, therefore, no questions are raised as to the independence of the basic postulates, the consistency or completeness of the system, or whether any logical errors have been made in the deduction of the theory's hypothesis. What will be examined is the extent to which the concepts and hypotheses of the market theories are related to observable phenomena.

1. Concepts and Conditions

If the concepts of a theory are devoid of empirical content then the hypotheses of which they are a part cannot be submitted to empirical test. While this may appear to be an unnecessarily obvious statement it emphasized the fact that the empirical content of a theory resides mainly in its concepts.

Consequently, even though the explanatory and predictive force of a concept can only be determined within the context of the relevant hypothesis or theory, its empirical content can be separately examined.^{1/}

A. The Concept of Demand

The first principal hypothesis of the theory of market behavior^{2/} is that a consumer's purchases of a specific item are determined by a function which is stated in terms of the price of that item, the prices of all other available products, and his income. This function is called the demand function. For a particular commodity, Q_j , and for a specific consumer, i , the function is represented by the relation:

$$D_{ij} = D_{ij}(p_1, p_2, \dots, p_j, \dots, p_m, y_i) \quad (j=1, 2, \dots, m) \quad (5.1)$$

where $(p_1, p_2, \dots, p_j, \dots, p_m)$ are the prices of the m commodities, and y_i is the i th consumer's income. If we are solely interested in the rate of change of the i th consumer's demand to a change in the price of Q_j then, as noted in the previous chapter, the demand function becomes a function of p_j alone, i.e.

$D_{ij} = D_{ij}(p_j)$. But what is the empirical interpretation of this concept and this relation? Clearly, in its general form, (5.1), there is no obvious connection between this function and observable phenomena. While the p 's represent prices

^{1/}For a detailed exposition of the problems associated with the empirical analysis of concepts in science see: C.G. Hempel, "The Fundamentals of Concept Formation in Empirical Science," International Encyclopedia of Unified Science, University of Chicago Press, Chicago, Vol. II, Chapter 7, 1955.

^{2/}This postulate is not just employed in the theory of a single commodity market. It is employed in a number of different market theories.

and y_1 represents income (5.1) does not specify directly the manner in which these variables are related. Consequently, in order to assess the empirical relevance of the concept of demand it is necessary to examine the basis from which this concept or function is derived. To facilitate this analysis attention will temporarily be restricted to the case where the consumer is confronted with a market that contains only two commodities.

The concept of demand emanates from an analysis of the choice behavior of an individual consumer during a specific interval of time. In a two commodity market the consumer's total purchases are a direct summation of the amount he spends on each of the two items. If one of the available commodities represents the consumer's stock of money,^{3/} then his total income for the period is equal to his total purchases, i.e. $y = p_1 q_1 + p_2 q_2$. But purchasing specific quantities of Q_1 and Q_2 the consumer derives a certain utility which is represented by the function $U = f(q_1, q_2)$. Since any combination of Q_1 and Q_2 will satisfy this function the equilibrium postulates are imposed upon the consumer's behavior to ensure that a unique selection is made. If a consumer's utility function is represented by the specific relation $U = q_1 q_2$ his purchases at equilibrium are determined in the following way:

Since the consumer's total income is given by, $y = p_1 q_1 + p_2 q_2$, and his utility from these purchases by, $U = q_1 q_2$, one can form with the aid of the Lagrangian multiplier λ a new expression which states that the consumer's utility is now given by: $W = q_1 q_2 + \lambda(y - p_1 q_1 - p_2 q_2)$. To find the equilibrium position of this utility function one applies the first order conditions as

^{3/}See pp. 54-55, Chapter 4.

follows: By taking partial differentials with respect to q_1 , q_2 , and λ , respectively and setting them equal to zero we get:

$$\frac{\partial W}{\partial q_1} = q_2 - p_1 \lambda = 0$$

$$\frac{\partial W}{\partial q_2} = q_1 - p_2 \lambda = 0 \quad (5.2)$$

$$\frac{\partial W}{\partial \lambda} = y - p_1 q_1 - p_2 q_2 = 0$$

Solving these equations for q_1 and q_2 ^{4/} two expressions result which relate q_1 and q_2 to their respective prices and the consumer's income:

$$q_1 = \frac{y}{2p_1} \quad q_2 = \frac{y}{2p_2} \quad (5.3)$$

Hence, under the condition that the consumer is maximizing his utility function during this period of time, the quantities of Q_1 and Q_2 he will purchase are strictly a function of his income and the price of the relevant item.^{5/}

The concept of demand is introduced by defining it to be equivalent to the quantity purchased of a particular commodity by a specific consumer at equilibrium. Accordingly, in this example, the demand for Q_1 is equal to the consumer's income divided by twice the market price of this commodity. Further, because the demand for Q_1 and Q_2 is solely a function of its price, a shift in the price will immediately alter the quantity demanded. To be more precise, the specific demand functions in this example allow us to conclude that an

^{4/} It is assumed here that the second order condition for a maximum is also satisfied.

^{5/} Note that (5.3) is a specific case of the general relation (4.4).

increase (decrease) in price will be followed by a decrease (increase) in the quantity demanded. As long as the demand functions are stated in this form it is clear that there is an inverse relation between market price and the quantity demanded.

Since a consumer's total income is assumed to be constant throughout the period in which these purchase decisions are made it is also apparent that the demand for Q_1 or Q_2 is homogeneous of degree zero. That is to say, if both prices and income were simultaneously reduced or increased by a similar percentage of their respective values, the quantity demanded would remain the same. Thus, as long as the consumer's income remains constant the demand function is a monotonically decreasing function of price. It represents the quantity a consumer will purchase of a specific commodity at the equilibrium point of his utility function.

Since the aggregate demand for a particular commodity is constructed by a summation of the quantity demanded by each individual, the aggregate demand function represents the total market purchases of a commodity only as long as each consumer is operating at his equilibrium point. If, for a particular period of time, the income of each consumer is constant, and if during this same period each consumer maximizes his utility function, then the aggregate demand function will also be a monotonically decreasing function of price alone.

It should be noted, however, that the decreasing monotonicity of the demand function depends upon the consumer's reaction to a price shift. If the price of Q_1 falls the effect on the consumer's purchases can be broken down into two components: the substitution and the income effect. To begin with, a drop in the price of Q_1 makes this commodity a better purchase relative to Q_2 . Accordingly,

as long as the consumer's utility function remains unchanged his purchases of Q_1 will increase. Thus, a fall in the price of Q_1 will induce the consumer to substitute more of Q_1 for Q_2 . But, at the same time the fall in the price of Q_1 increases the consumer's total income, i.e. it increases the total amount of commodities the consumer can purchase.

For example, before the price of Q_1 is altered the consumer's total income is given by, $y = p_1q_1 + p_2q_2$. If the consumer is maximizing his utility his position on his utility curve is at the point of tangency between it and his income line. In Figure 1 this point is represented by A where U_a represents his current level of utility and $Y-Y$ his income line. When the price of Q_1 falls the demand for Q_1 increases. And if the consumer remains on the same utility function

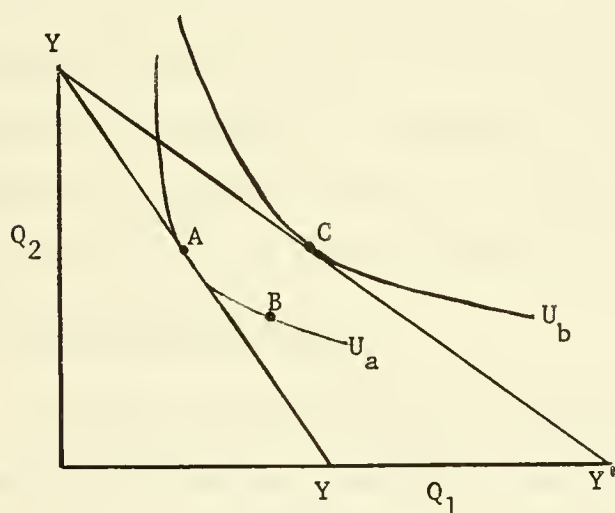


Figure 1

the increase in consumption of Q_1 can be represented by point B. But, unless the consumer purchases a sufficiently large amount of Q_1 so that the amount he now spends on Q_1 is equal to the original amount spent on Q_1 , i.e., p_1q_1 , he will also be able to purchase more of Q_2 . This extra quantity that he can purchase of Q_2 represents the effect the

price change has on his real income. If due to the decrease in price the consumer decides to consume at point C it is clear that not only has he been able to shift to a higher level of utility, U_b , with a corresponding shift in his income line $Y-Y'$, but this shift has allowed him to consume more of both Q_1 and Q_2 .

However, it is not always the case that the consumer will purchase more of both Q_1 and Q_2 . If Q_1 is an inferior good then a decrease in its price will not lead to an increase in its consumption. On the contrary, in this case the income effect will dominate the substitution effect and the consumer will spend the extra amount on Q_2 . Accordingly if after a fall in the price of a particular commodity the income effect is sufficient to offset the substitution effect the demand function for that commodity cannot be represented as a monotonically decreasing function of price. If all consumers behave in this fashion with respect to this specific commodity then the aggregate demand function can no longer be represented as having a negative slope. Moreover, if a particular item is an inferior good only to a certain number of the consumers in the market, then to determine the slope of the aggregate demand function it is necessary to know the relative magnitude of the total substitution and income effects. Since the substitution effect always increases the demand for a commodity it is only the presence of a large and positive income effect which will allow the fall in a price to effectively decrease the demand for that commodity.^{6/}

B. The Concept of Supply

The second main hypothesis of the theory of market behavior states that the amount of a specific item which a firm will produce is a function of the market price alone. The supply function itself is derived by defining it to be identical to the relevant portion of the firm's marginal cost curve. Under short-run

^{6/}For a more detailed description of the theory of consumer behavior see: P.A. Samuelson, Foundations of Economic Analysis, Harvard University Press, 1947, Chapter 5; and G.P.E. Clarkson, op. cit., Chapter 3.

conditions the supply function is defined as that, part of the marginal cost curve which lies above the firm's average variable cost curve. Under long-run conditions it is identical to the segment of the marginal cost curve that lies above the average cost curve. As a result, to assess the empirical relevance of the supply function one must in turn inspect the basis from which the marginal cost curve is developed. Cost function is developed from a knowledge of the firm's production function, a function which relates the cost of the variable and fixed inputs to the production process, and a function which describes the way in which the inputs should be increased if the firm's output is to be expanded at a minimum of cost. By combining these three functions a single relation is produced which is an explicit function of the levels of output and the amount of fixed cost. This is the total cost function and it represents the minimum cost at which each level of output can be produced by this particular firm. Once the firm's cost function is determined the marginal cost function is derived by taking the first derivative of the cost function with respect to output. Since the cost relation is a function of variable and fixed costs, the marginal cost relation is a function of variable cost alone.

For example, if the total cost relation of a particular firm for a product q can be represented by $C = \Phi(q) + b$, where b represents the fixed cost associated with producing q , then the marginal cost of producing q is given by:

$$\frac{d}{dq} C = \frac{d}{dq} \Phi(q) \quad (5.4)$$

where $\Phi(q)$ represents the variable costs associated with different levels of output. Since the cost relation, C , gives the minimum cost at which each level of output can be produced, the marginal cost curve gives the minimum, additional, variable cost incurred at each level of output.

Now the supply function is identical to that part of the marginal cost curve which lies above the average variable cost function in the short-run and the average cost function in the long-run. Hence to determine the beginning of the supply function--that is, the point below which the firm will not produce any output--one needs to locate the intersection of the average variable and long-run average cost functions with the marginal cost function. Since the average variable cost function is given by $\frac{\Phi(q)}{q}$ it is easy to determine the point of intersection between it and the marginal cost function given by (5.4). The intersection takes place at the minimum point of the average variable cost function. In the same manner it can be shown that in the long-run when all costs are variable costs the intersection takes place at the minimum point of the average cost function.

Once the beginning of the supply curve is determined its only remaining important characteristic is its slope. To determine the slope of the supply curve we need to identify the slope of the marginal cost curve. This is accomplished in the following way: Consider a firm which is selling its output, q , at the current market price, p . The firm's revenue is given by the quantity sold multiplied by the price. Profit is the difference between the revenue and the cost of production. Hence, the firm's profit can be represented by:

$$\Pi = pq - \Phi(q) - b \quad (5.5)$$

To determine the output the firm will produce the first-order condition is applied--i.e., that profit must be at a maximum--and the first derivative of profit is taken with respect to output.

$$\frac{d\pi}{dq} = p - \frac{d}{dq} \Phi(q) = 0 \quad \underline{7/} \quad (5.6)$$

or

$$p = \frac{d}{dq} \Phi(q) = \text{Marginal Cost} \quad (5.7)$$

Consequently, the firm produces at the point where marginal cost equals the market price. To find the slope of the marginal cost curve the second-order for a maximum is applied, i.e., $\frac{d^2\pi}{dq^2} < 0$, to obtain:

$$\frac{d^2\pi}{dq^2} = -\frac{d^2}{dq^2} \Phi(q) < 0 \quad (5.8)$$

Hence, the slope of the marginal cost curve which is given by, $\frac{d}{dq} [\frac{d}{dq} \Phi(q)] > 0$, is greater than zero. Accordingly, the supply function for a particular firm has a positive slope and is a monotonically increasing function of price alone. Since, the aggregate supply function for a market is a direct summation of the individual functions, the aggregate supply function has the same characteristics as long as all firms have positively sloped supply functions.

As has already been noted, there is one situation--i.e. where external economics or diseconomics occur--in which the aggregate supply function may be negatively sloped.^{8/} This is the case where the cost function of each firm is no longer independent of the output levels of other firms but is instead dependent upon such outputs. If these dependencies are such that the relevant portion of the firm's marginal cost curve becomes negatively sloped then the supply function will also have a negative slope. As a result, unless the supply

^{7/}Note the similarity between (5.5) and (5.6), and their general formulation in (4.11) and (4.12).

^{8/}See section 2 Chapter 4.

function of each firm has the same slope, or unless the relative magnitude of the respective slopes is known, it is no longer possible to determine the slope of the aggregate supply function.

C. The Concept of Equilibrium

A market is at equilibrium if the aggregate amount demanded equals the aggregate amount supplied. If the market is at equilibrium then there is only one price for each product in the market. Each consumer is at equilibrium if with each set of purchases he is maximizing his utility function. Similarly each producer is at equilibrium if his output is such that he is operating at the point where his marginal cost function equals the market price. If the market conditions are such that there are no external economics or diseconomics, and if the income effect of price changes can be ignored, then from a knowledge of producers' cost functions and consumers' demand functions the equilibrium price at which each commodity will be bought and sold can be determined. Further from a knowledge of the slopes of the aggregate supply and demand functions it is possible to determine whether the equilibrium is stable or not. Moreover from the equilibrium positions of consumers and producers the theory provides a set of relations which must hold if market equilibrium is to be attained. Consequently, under the full set of equilibrium conditions the theory of market behavior provides a set of hypotheses with which all market transactions and behavior can be determined.

2. Testing the Theory's Hypotheses

In order to submit a theory to empirical test there must be at least one hypotheses, or consequence of a hypothesis, that under appropriate initial

conditions refers directly to observable phenomena. If the theory does not contain such a hypotheses then, as noted in Chapter 2, the theory cannot be employed to explain or predict the occurrence of observable events. Thus, while the theory may remain as an interesting deductive system, it is not possible to consider it as a part of empirical science.

Since we, as economists, are interested in being able to employ the theory of market equilibrium to explain and predict market behavior we must first make sure that the theory contains at least one testable hypotheses. If the theory contains such a testable hypothesis then we can proceed to employ market data to check, test, and amend the theory. However, if the theory does not contain a testable hypothesis then we cannot employ it an empirical theory and must classify it as an uninterpreted deductive system.

To subject the theory or any of its hypotheses to empirical test presents the experimenter with several problems. The first main obstacle is the manner in which the theory is to be considered. If the theory is to be tested directly against observed market behavior then it is necessary to determine, before a test is conducted, that the initial conditions are all satisfied.

For example, to be able to test any of the hypotheses concerning consumer behavior one must first determine that each consumer is maximizing his utility function. Unless this condition is satisfied it is not possible to test these hypotheses. For the theory does not hold except under equilibrium conditions. Consequently, the first task is to examine the behavior of certain consumers and ascertain whether they are behaving so as to maximize their utility functions. To carry out such an investigation requires one to be able to identify a consumer's utility surface and simultaneously determine whether he is situated at the maximum point on this surface. Since it is not possible to employ one set of

observations to simultaneously determine both the utility surface and whether the consumer is at a maximum of utility, the best that can be done is to take observations at succeeding intervals of time. But the minute the observations are extended over several time periods a further complication is introduced.

The consumer's utility function, and consequently his demand function as well, is defined only over a single interval of time. For tastes and income must be held constant to permit the hypotheses about consumer behavior to be inferred. Accordingly, the theory allows the consumer to shift to a new equilibrium position at the beginning of each interval of time. Hence, unless one has a method whereby one can determine that tastes and income have not changed between intervals one cannot employ the data from two different periods of time to substantiate one utility surface. Clearly, if the constancy of tastes and preferences were separately measurable for a consumer over time, then one could employ this knowledge to establish the hypotheses governing his choice behavior. But, the theory requires that tastes remain constant within each interval without providing a basis from which this condition can be tested. Thus, it is manifestly not possible to determine from one set of observations whether this condition has been fulfilled.

Unless this condition can be independently established the hypotheses about consumer behavior cannot be submitted to test. That this conclusion must hold follows directly from the earlier analysis of the conditions under which a hypothesis or theory can be tested. If the initial conditions under which the theory is supposed to hold are represented by, $\underline{P}$, and the relevant hypotheses by, $\underline{Q}$, then the theory can be represented by the conditional statement, if $\underline{P}$ then $\underline{Q}$, i.e. $\underline{P} \rightarrow \underline{Q}$. To subject this relation to test it must be possible to disconfirm it.

The only condition under which this is so is when there is evidence supporting the propositions included in P. If the empirical truth value of P is unknown it is not possible to determine the empirical truth value of the relation $\underline{P} \rightarrow \underline{Q}$. While evidence supporting Q may or may not be easy to find, such evidence cannot by itself corroborate the entire relation. Consequently, before it even makes sense to inquire whether a particular set of consumers are maximizing their utility it must first be possible to measure, by an independent set of tests, the constancy of their preferences.

At the same time consider the problem of testing for the slope of the aggregate demand function. The theory asserts that the aggregate demand function will have a negative slope if consumers are maximizing their utility functions, if the substitution effect is always greater than the income effect, and if we are not dealing with commodities, such as a number of luxury items, for which the individual demand functions are positively sloped. Under these conditions the aggregate demand function will have a negative slope. Manifestly, to test this hypothesis one must first be able to empirically determine whether in a particular market situation these conditions are satisfied. As long as the commodities are superior goods the substitution effect will be sufficiently larger than the income effect to satisfy this requirement. But how does one identify, prior to and independently of a particular investigation, which goods for a specific set of consumers are superior? If there were a set of tests which were always able to identify inferior goods relative to a particular group of consumers the first of these three conditions could be established. But the theory does not provide a mechanism for establishing such a set of tests. The only way the theory can be used to classify commodities is to observe, in a particular instance, whether the

demand function after a price change for an individual or group of consumers is positively or negatively sloped. The same comments hold for the third condition. For the theory does not provide any method for independently establishing the slope of the demand function for cases when it is not negatively sloped. Since the impossibility of testing the second condition has already been discussed it is clear that it is not possible to independently determine whether these conditions are fulfilled. To be unable to empirically establish the initial conditions implies, of course, that one cannot submit the hypotheses to empirical test.^{9/}

If the initial conditions on the demand side of the theory perhaps cannot be empirically established perhaps it is possible to do so on the supply side. For example, in order to test the hypothesis that the supply function of a particular firm is positively sloped all one needs to be able to do is to determine the slope of the firm's marginal cost function. The marginal cost function is directly derived from the firm's total cost function. Thus, the first problem is to empirically establish the nature of this function for the firm in question. But, to determine the total cost function of a firm it is not sufficient merely to discover a relation which yields the total cost to the firm of producing a certain output.

The total cost function of the theory defines a relation between the cost of inputs and outputs such that this is the minimum cost at which such an output can be produced by this firm. As is noted above, the cost function is derived from the firm's production function, its minimum cost relation, and a function specifying

^{9/}A more detailed analysis of the empirical content of the classical theory of consumer demand is to be found in: G.P.E. Clarkson, op. cit., Chapters 4, 5, and 6.

the manner in which inputs should be increased if output is increased so as to remain at a minimum of cost. Manifestly, if a firm's cost function is to be employed as the basis for the supply function, one must first make sure that the firm is producing at a point of minimum cost. But how does one ensure that this condition is satisfied? Further, not only is it necessary to determine that the firm is operating at a point of minimum cost, but it must also be shown that it has set its level of output such that marginal cost equals price. If this latter condition is not satisfied, then for obvious reasons it is not possible to inspect the properties of the supply function.

In an empirical investigation of a firm's behavior one can readily examine, for a particular period of time, the cost of its inputs, the current level of fixed cost, the amount of output product, as well as the price per unit received for this output. Clearly, this examination can be carried out in great detail so as to develop an accurate picture of the cost structure of the firm. But, this is not enough! Along with these data it must also be possible to tell whether the firm's production process is such that it is operating at a point of minimum cost. While observations can provide data on current costs they cannot at the same time provide any information on whether this is a minimum level or not. To determine whether the firm is at a minimum one would need to be able to assess all the possible ways of combining the inputs to achieve the same level of output. But the actual observations cannot yield this information at the same time as they are depicting the firm's current behavior. Once again a separate and independent means is required for checking the empirical truth value of these initial conditions.

Similar difficulties are encountered if one is dealing with aggregate data and wish to test the hypothesis that the aggregate supply function has a positive slope. In this case, in addition to the initial conditions which pertain to the

determination of an individual supply function, it must be possible to identify whether external economies or diseconomies exist in the market. Unless one can determine when, for example, external economies are present prior to and independently of an inspection of the slope of a supply function, then it is not possible to conduct any empirically meaningful tests on the slope of an aggregate supply function.

For example, consider the analysis of the effect of externalities upon the slope of the supply function discussed in the previous chapter. Here the analysis began by supposing that one could write down the profit functions for all n firms in the market. These functions are represented by:

$$\Pi_i = pq_i - C_i \quad (i=1,2,\dots,n)^{10/} \quad (5.9)$$

where C_i represents the total cost function of the i th firm which is dependent upon the level of output of all other firms, i.e. $C_i = \Phi_i(q_1, q_2, \dots, q_n)$. An equilibrium solution for this market is described by taking the partial differentials of each firm's profit function with respect to the commodity it produces, setting these relations equal to zero and solving for the set of equilibrium outputs, q_i^0 . The supply functions are derived from the equilibrium solution. If their slopes are known the slope of the aggregate function can be deduced. But, due to the presence of externalities one can only determine the slope of an individual supply function if one already knows the signs of all the coefficients in its cost function. It is not sufficient to know that each firm is maximizing its net revenue. It must also be possible to ascertain the effect of each firm's output decision on that of every other firm. Since the cost function of a firm is

^{10/}See (4.11) and the accompanying discussion in Chapter 4.

empirically non-determinable when externalities are absent, the same reasons preclude the measurement of the values of the relevant coefficients when externalities are present.

In the analysis of externalities in Chapter 4 it is noted that if the values of the coefficients are not known in the individual cost functions, then the theory still permits one to solve for the slopes of the supply functions by an application of the technique of comparative statics.^{11/} To employ this method of analysis one must first derive the equilibrium solution for the total market. Employing this solution as the initial conditions a new set of relations are introduced by taking the partial differentials of the equilibrium relations with respect to the outputs of the individual firms. The process of partial differentiation requires, of course, all other parameters and variables to be treated as constants. Accordingly, the rates of change that are derived by this method only hold under conditions where all other factors can be shown to have remained constant. Further, comparative statics requires the initial conditions to be the equilibrium solution of the market. Therefore, since it is not possible to empirically determine when the market is at equilibrium it follows that it is also not possible to test for the empirical significance of the rates of change generated by this method.

In the previous sentence it is asserted that it is not possible to determine when a market is at equilibrium. While I have demonstrated that it is not possible to submit the hypotheses of the theory of market behavior to a process of refutation by empirical test, it may well be claimed by some that they can at least tell when a market is in equilibrium. After all, a market is in equilibrium when

^{11/} See pp. 59-62.

aggregate demand equals aggregate supply. And it follows that this condition is met when there is only one price for each commodity on the market. Consequently, all that needs to be done, it might be argued, is to find a market in which one price has prevailed for each commodity over some reasonable interval of time and this will be an example of a commodity market in competitive equilibrium.

Unfortunately, however, it is not legitimate to infer from the observation of one price, during a reasonable interval of time, to the presence of an equilibrium position in the market. The market may have only one price for each commodity for a large number of reasons, e.g., a variety of collusive business practices, or government intervention in pricing decisions, but only one of them ensures that the market is at the theoretical equilibrium. This is, of course, the requirement that each firm is maximizing his net revenue and each consumer is maximizing his utility function. Since it is not possible to determine by observation when this requirement is met, it is clearly not possible to test whether the one price per commodity prevailing in the market is an indication of equilibrium or not. If an empirical investigation cannot be establish when a market is in equilibrium then none of the hypotheses which employ market equilibrium as an initial condition can be confuted by empirical test.^{12/}

Consider, for example, the case of the spatially distributed firms dealing in dairy products which is examined in some detail in the last chapter.^{13/} In this case there are two sets of firms. The first set are nearer the central

^{12/} The problem of testing for the empirical validity of equilibrium conditions is further discussed in the next chapter and again in Chapter 11. A detailed examination of this issue with respect to decision behavior of individuals and groups is presented in: G.P.E. Clarkson, op. cit., Chapter 5.

^{13/} See pp. 66-70.

market and have a transportation cost of $\alpha_1=10$ dollars per unit. While the second set incur transportation costs of $\alpha_2=13$ dollars per unit. Aside from this difference all firms are supposed to have identical cost functions. If all firms have identical cost functions then it follows that they must have identical production functions, cost relations, and expansion functions. Now it is clear that one can examine the cost structure of each firm and determine the process by which it transforms inputs into outputs. But even if under the most detailed scrutiny, each firm has identical production processes and cost structures, one is still unable to empirically ascertain whether this cost structure represents the minimum attainable. While the presumptive evidence might be strongly in favor of such a conclusion, there is no independent measure by which its empirical validity can be determined. Unless this cost function represents the minimum attainable with the current technology then the first condition for a market equilibrium has not been met.

Continuing with the example, the next step in the analysis is to derive the supply function for each class of firms. This is accomplished by finding that output which maximizes their net profit. By applying the first-order condition for equilibrium to the profit function and solving for the quantity produced one can analytically specify the supply function for each class of firm. However, to subject the results of this analysis to empirical test one must once again be able to demonstrate that the initial conditions are empirically true. This implies establishing the fact that these firms are operating at a position of maximum net revenue. But, even to observe that all firms within each class are producing the same output is not sufficient evidence to guarantee they are maximizing their net revenue. While such behavior would clearly be consistent with the theory, it cannot be employed to corroborate the theory's conclusions.

To do so requires an independent check on the empirical validity of the initial conditions. Moreover, to determine that the initial conditions are satisfied requires an independent measure of when a firm is operating at a maximum of net revenue. Since the theory does not provide the interpretive rules whereby such measurements can be made it is manifestly not possible to subject these hypotheses to test.

Summary and Conclusions

In the first part of this chapter the conditions under which the concepts of demand, supply, and market equilibrium are derived and employed are inspected in some detail. The theoretical basis of each concept is examined as well as the criteria by which it is possible to measure such observable attributes as slope, output, and price. At the same time it is noted that to subject a theory or any of its hypotheses to empirical test the initial conditions must be shown to be empirically true.

The remainder of the chapter is then devoted to an analysis of some of the main hypotheses of the theory of market equilibrium. The object of this examination is to discover whether the theory contains any hypotheses which can be subjected to test. In carrying out this investigation it is shown that all the theory's hypotheses have at least one equilibrium requirement as part of their initial conditions. Further, it is demonstrated that within the context of the theory it is not possible in any specific case to empirically determine whether such equilibrium conditions are satisfied. If the initial conditions cannot be shown to hold the hypothesis cannot be subjected to test. It follows, therefore, that none of the theory's hypotheses can be subjected to a process of refutation by

empirical test. Consequently, one is forced to conclude that the theory of market equilibrium is devoid of empirical content. It can make no claim to refer to observable phenomena. This is a strong and serious conclusion, and the next chapter is devoted to examining the implications of this result for classical theories of microeconomic behavior.

Chapter 6

AN EMPIRICAL ANALYSIS OF THE CLASSICAL DEDUCTIVE SYSTEM

In order to explain or predict the occurrence of an economic event it is necessary to have an economic theory that can satisfy the following two conditions: (i) the theory must contain at least one hypothesis which can be directly submitted to empirical test; (ii) such hypotheses must have been submitted to and have survived at least one test. The theory of market equilibrium, however, is unable to satisfy the first requirement. Accordingly, it is clear that it is not possible to employ this theory to explain or predict the occurrence of the market phenomena to which it refers. This is an important conclusion. It not only implies that the theory is empirically vacuous, but it also suggests that the obstacle to empirical interpretation lies within the deductive system from which the theory is developed. If this latter inference is correct, then all theories which are based upon that same deductive system will encounter similar empirical difficulties. That is to say, if the absence of empirical content can be shown to be a result of the way in which classical theories of economics are developed then it follows that none of these theories will contain empirical hypotheses which refer to observable economic behavior. If such is the case then none of these theories can be employed to explain or predict the occurrence of economic events. The seriousness of this corollary warrants a detailed investigation of its validity, and this chapter is devoted to such an examination.

1. The Basic Deductive System

In Chapter 3 the examination of the classical foundations of economic analysis began with an example of how a firm's reaction to a tax upon its output is determined. By an application of basic deductive system it is shown that there is a negative relation between the rate of change of the firm's output and the tax rate--i.e. if the tax rate is increased the firm's output will decrease and vice versa. Since it is perfectly clear that one can observe both an increase in a firm's taxes as well as a decrease in its level of output this instance appears to present a counter example to the argument of the previous chapter. Consequently, it is reasonable to begin the investigation of the classical deductive system by a re-examination of the empirical content of this example.

In this case a firm is considered for which it is supposed that the demand curve for its output, $x_p(x)$ is already known. To simplify the analysis the firm produced only one item, X . (If a firm was selected which produced many items then the firm's demand curve would have to be defined in terms of all such items.) Furthermore it is supposed that there is sufficient information on the firm's production process so that we knew the relation between the total production cost for the firm and its output, $C(x)$ is also known. With these two functions it is then possible to specify the profit function for the firm, in the normal manner, as the difference between its total revenue and the total cost of producing a certain output at a specific price:

$$\Pi = x_p(x) - C(x)$$

A tax on output is then imposed upon the firm and is included as a further item in the profit function:

$$\Pi = x_p(x) - C(x) - t(x) \tag{6.1}$$

In order to determine the effect of the tax rate on output one first has to derive the expression which specifies the equilibrium relation between output and the tax rate. To generate the equilibrium solution the first-order condition for a maximum with respect to output is applied to (6.1) which yields:

$$t = \frac{\partial}{\partial x} [xp(x) - C(x)] \quad (6.2)$$

To ensure that this is a position of maximum net revenue the second-order condition for a maximum must also be satisfied, i.e.:

$$\frac{\partial^2}{\partial x^2} [xp(x) - C(x)] < 0 \quad (6.3)$$

But before it is empirically meaningful to apply these equilibrium conditions to (6.1) production cost function, C(x), represents for this firm the lowest total production cost at which each level of output is produced. Unless the cost function has this property it makes no sense to apply the equilibrium conditions and solve for the equilibrium level of output with respect to the tax rate as in (6.2). Moreover, unless the demand function, xp(x), can be shown to represent the demand at various prices for this firm's product it makes no empirical sense to construct the profit function (6.1) in this manner.

In the previous chapter it is argued that by taking observations at any one point in time it is not possible to determine whether a firm is operating at a point of minimum cost. If several observations are taken over succeeding intervals of time at varying levels of output then it is necessary to be able to measure the minimum production cost at each of these output levels. A minimum can only be ascertained if all possible combinations of inputs and their respective costs are measured against a specific level of output. Further, the production cost function represents the locus of these minimum points as output is varied. Thus,

to determine the production cost function one needs to be able to observe and ascertain whether or not the firm is always operating at a minimum of cost. Since at any given point in time the theory does not provide us with sufficient interpretive rules to allow us to measure whether the firm is at a minimum of cost, it is manifestly not possible to empirically determine the locus of a series of such points. If the function, $C(x)$ is not empirically specified, then it is not possible to specify the profit relation given in (6.1). Consequently, for any particular firm it follows that one is unable to demonstrate that the function represented by (6.2) is the actual equilibrium solution.

To ensure the presence of a maximum the second-order condition, (6.3), must be satisfied. But, this requires an empirical knowledge of the functions $x_p(x)$ and $C(x)$ such their second partial derivatives can be evaluated with respect to output. If it is not possible to empirically determine the minimum production cost function of a firm it is certainly not feasible to evaluate its second partial derivative with respect to output.

In Chapter 3 the second condition is employed to infer the direction of the equilibrium rate of change of output with respect to the tax rate, $\left(\frac{\partial x}{\partial t}\right)^0$. Unfortunately, one cannot determine for a specific firm whether the second-order condition is empirically satisfied. Consequently, the derivation of the rate of change output with respect to a change in the tax rate results in a situation where once again it is not possible to determine if the initial conditions are empirically true.

It could be argued, however, that the equilibrium method of solution only directly applies to the case of an ideal firm, market, or consumer, as the case may be. In this respect the above example would yield the direction of the

equilibrium rate of change of output with respect to the tax rate for an ideal firm that operated at the relevant equilibrium position. Further, it would be pointed out that most of the theories in the physical sciences are formulated in this manner in that they only refer to specific, ideal cases. For instance, the theory about the behavior of gases states that the pressure of a gas is a function of its temperature and volume. This theory is defined and the function's parameters are specified in terms of an ideal gas--namely, a specific type of gas which has a number of idealized properties. To relate such an ideal theory to specific, observable cases interpretive rules are provided which permit the experimenter to empirically establish the presence or absence of the requisite initial conditions. Since the initial conditions can be observed and the relevant pressure, temperature, and volume can be measured, the theory itself can be submitted to test in a variety of specific cases.^{1/} Accordingly, if one is to treat economic theory in a similar manner the theory itself must provide sufficient interpretive rules to permit part or all of the theory to be confronted by empirical test.

The analysis in the last chapter argues, in effect, that economic theory does not contain such interpretive rules. That is to say, it is demonstrated that the deductive system underlying economic theory is such that all hypotheses require

^{1/}For an extensive analysis of the uses and misuses of the concept of "ideal" type, case, and theory in empirical science see: E. Nagel and Carl G. Hempel, "Symposium: Problems of Concept and Theory Formation in the Social Sciences," Science, Language, and Human Rights, American Philosophical Association, Eastern Division, University of Pittsburgh Press, Philadelphia, 1952, Vol. I, pp. 43-86. (Reprinted as: E. Nagel, Problems of Concept and Theory Formation in the Social Sciences," and C.G. Hempel, "Typological Methods in the Social Sciences," in M. Natanson, (ed), Philosophy of the Social Sciences, Random House, New York, 1963, pp. 189-209, and pp. 210-230, respectively.

some equilibrium conditions to be met as part of their initial conditions. Since economic theory does not provide any interpretive rules by which these initial conditions can be observed it is concluded that it is not possible to subject the theory to empirical test. In order to subject both this claim and its corollary to a more detailed scrutiny it is necessary to re-examine the deductive system upon which all classical economic theory is based.

2. The General Deductive System

In the general case the economic system is represented by $\underline{n}$ functional relations each of which contains $\underline{n}$ variables and $\underline{m}$ parameters. If each functional relation is represented by the notation:

$$f(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) = 0$$

then the total system of functions is represented by:

$$f^i(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) = 0 \quad (i=1, 2, \dots, n) \quad (6.4)$$

If this system is to represent the basis of a deductive system then the $\underline{n}$ relations in (6.4) must be independent of each other as well as consistent with each other. While both these conditions must be satisfied by any specific application, neither depend upon the empirical content of the system. Consequently, for the purposes of this analysis the process by which one determines whether these criteria are met will be ignored. Instead suppose that they are in fact satisfied.^{2/}

The basis of the economic system, then, is represented by (6.4). The

^{2/}The criteria of independence and consistency in deductive systems, i.e. calculi, are discussed in detail in: A. Church, op. cit., Chapter 1.

equilibrium solution is derived by applying the following constraints: First, a set of initial conditions are chosen--namely, a specific set of values are given to the parameters $(\alpha_1, \alpha_2, \dots, \alpha_m)$ for each of the $\underline{n}$ functional relations. Second, the first-order condition

$$\frac{\partial}{\partial x_i} f^i(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) = 0 \quad (6.5)$$

is imposed and a further set of $\underline{n}$ relations represented by (6.5) are generated. Since the equilibrium point represents a maximum the third step consists of examining whether the second-order condition

$$\frac{\partial^2}{\partial x_i^2} f^i(x_1, x_2, \dots, x_n, \alpha_1, \alpha_2, \dots, \alpha_m) < 0 \quad (6.6)$$

is satisfied for all $\underline{n}$ relations in (6.5). If each of these conditions is imposed and if each is satisfied then the equilibrium solution to (6.4) is given by a set of $\underline{n}$ values of the variables $\underline{x}_i$. These values are stated in terms of the initial values given to the parameters $(\alpha_1, \alpha_2, \dots, \alpha_m)$ and can be represented in functional form as:

$$x_i^0 = g^i(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0) \quad (i=1, 2, \dots, n)^{3/} \quad (6.7)$$

By this deductive process a specific set of values are arrived at for the variables $\underline{x}_i^0$ in terms of the initial values given to the parameters $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$. Now, if within the context of a particular case, it were possible to observe and test the relations given in (6.5) it would imply that it was also possible to observe and record the appropriate initial values of the parameters.

^{3/} Relations (6.5), (6.6), and (6.7) are, of course, identical to (3.9), (3.10), and (3.8) respectively.

Under such conditions it might then be possible to observe some of the values of the $\underline{x}_i$ and as a consequence test whether the system under consideration is at equilibrium.

In a specific application of this deductive process to an economic system empirical difficulties are encountered at all three stages. To begin with the functions represented by (6.5) are themselves equilibrium relations. In the example mentioned in this chapter they would represent the production and cost functions of $\underline{n}$ firms whose output was or is to be taxed. Indeed, the relevant production and cost functions are those which specify the relation between inputs and output such that the output is being produced at a minimum of cost. In the case of a consumer the functions in (6.5) represent the relevant demand relations. These relations are also equilibrium functions and depend upon the continued maximization by each consumer of his utility function. Since, the relations in (6.5) are equilibrium relations this implies that the initial parameter values $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ are also equilibrium values. Accordingly, to empirically determine their values one would need to go back a step and inspect the process by which they were generated. Such an analysis, however, leads to exactly the same position as was examined in the last chapter. There it was noted that it was not possible to employ the results of a single empirical investigation to determine whether a firm was employing a minimum cost and production schedule or whether a consumer was maximizing his utility. Thus, as one cannot submit the process by which the equilibrium parameters are generated to test, it is clear that there is not an empirical process by which one can establish the initial values $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$. Without such a process it is not possible to empirically determine which parameter values satisfy the first of the three constraints.

The third constraint given by (6.6) poses another serious obstacle. If one is unable to empirically establish the equilibrium values of $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ how can one possibly compute whether the second-order partial differential equations have values less than zero? Since not all relations one might care to write down will satisfy (6.4), and since one is unable to ascertain the relevant parameter values, it follows that one is also unable to determine for any specific case whether (6.6) is satisfied or not.

The same remarks apply with equal force to the second constraint represented by (6.5) as well as to the solution to these relations given by (7). If the values of $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ cannot be empirically established how is one to test for the equilibrium values of the variables, $x_i^0 = g^i(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$? Manifestly, one is in the same position as before where the inability to establish the initial conditions precludes the possibility of being able to subject the resulting relations to empirical test.

In order to circumvent the full force of these obstacles to empirical interpretation a method was devised which permits the determination of the directional change of individual variables in response to selected changes in the initial conditions. This method of comparative statics begins, as already described, by first assuming that it is possible to solve for the equilibrium solution given by (6.7). Once the relations in (6.7) are known selected shifts in the values of $(\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ are made so as to determine the effect these shifts have upon the direction of the rate of change of certain variables. The deductive process proceeds by taking the first partial derivative of the equilibrium relations $f^i(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$ with respect to one of the

parameter values, say α_1 . The result of this operation is a system of $\underline{n}$ partial differential relations where all parameters and variables except the ones being differentiated, are treated as constants. This system of relations is represented by:

$$f_{x_1}^i \left(\frac{\partial x_1}{\partial \alpha_1} \right)^0 + f_{x_2}^i \left(\frac{\partial x_2}{\partial \alpha_1} \right)^0 + \dots + f_{x_n}^i \left(\frac{\partial x_n}{\partial \alpha_1} \right)^0 = -f_{\alpha_1}^i \quad (6.8)$$

where

$$f_{x_j}^i = \frac{\partial f^i}{\partial x_j} (x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$$

and

$$f_{\alpha_1}^i = \frac{\partial f^i}{\partial \alpha_1} (x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$$

Since all other variables and parameters are treated as constants, during the process of partial differentiation, all the $f_{\alpha_1}^i$ terms represent coefficients of the variables $\left(\frac{\partial x_i}{\partial \alpha_1} \right)^0$. Hence, as noted in Chapter 3, (6.8) represents a system of $\underline{n}$ linear equations in $\underline{n}$ unknowns. The unknowns in turn represent the equilibrium rates of change with respect to the shift in the value of the parameter α_1 .

The desired result is to be able to determine the sign of each of the variables $\left(\frac{\partial x_i}{\partial \alpha_1} \right)^0$. In other words, the procedure is designed to ascertain the sign of the coefficients given by

$$f_{x_j}^i = \frac{\partial f^i}{\partial x_j} (x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$$

The analytical process by which these results are derived is described in Chapter 3 and need not be repeated here,^{4/} except to note that the resolution of the signs

^{4/} See pp. 40-47.

requires the application of a number of further conditions. These constraints make use of the equilibrium conditions as well as the special requirement that a shift in one parameter may only affect one of the relations in (6.8). The principal result of this analysis is a criterion function with which one is able to determine the sign of the variables $(\frac{\partial x_i}{\partial \alpha_1})$. This criterion function is given by

$$f_{x_i \alpha_i} \left(\frac{\partial x_i}{\partial \alpha_1} \right) > 0 \quad (6.9)$$

where

$$f_{x_i \alpha_i} = \left(\frac{\partial}{\partial x_i} \right) \left(\frac{\partial}{\partial \alpha_i} \right) f(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0) \quad (i=1, 2, \dots, n)$$

Disregarding, for the moment, the special conditions under which this criterion is developed one can now inspect it to see whether it permits the direction of these rates of change to be empirically established. If one is to be able to test the result of an application of this criterion one must be able to specify the function represented by, $f(x_1^0, x_2^0, \dots, x_n^0, \alpha_1^0, \alpha_2^0, \dots, \alpha_m^0)$. But this is the equilibrium relation derived from the solution of (6.5). But, one cannot empirically establish the solution to (6.5). Hence, it is hard to know how to establish the empirical relevance of the criterion in (6.9). Indeed, to determine the empirical content of (6.9) it is necessary to be able to demonstrate the empirical relevance of the other special conditions. That is to say, it must be defended upon observational grounds that not only is it possible to treat a shift in one parameter at a time, but it is also possible to show that the effects of this shift are restricted to one of the equilibrium relations. Moreover, the theory does not permit the empirical specification of the equilibrium relations, let alone their initial conditions. Thus, the empirical relevance of these constraints is meagre to the vanishing point.

Since the technique of comparative statics is based upon the equilibrium solution to the original set of relations (6.4) all hypotheses which are deduced by this technique have some equilibrium values as their initial conditions. It follows, therefore, that all propositions or hypotheses produced by this deductive system have as a part of their initial conditions some unobservable equilibrium values.

To subject an hypothesis to test it must be possible to observe the occurrence of its initial conditions. To be a part of empirical science a theory must contain at least one hypothesis that can be submitted to the process of refutation by empirical test. But the deductive system of classical economics by relying upon equilibrium constraints precludes the possibility of generating testable hypotheses. Consequently, it is the deductive system itself which confers the empirical vacuity upon the hypotheses and theory of classical economics.

3. The Market Conditions

So far the analysis of the deductive system has not included the empirical conditions delimiting the type of market under which these theories are supposed to hold. Clearly before a market theory could be submitted to test one would need to be able to show that the requisite initial conditions were satisfied by the specific market under investigation. For example, if one were testing the theory of market behavior under perfect competition one would need to empirically establish whether: (i) All firms within the market are producing a homogeneous product. A market for dairy products is a reasonable example of such a market and was employed in Chapter 4. But, if brand names and other promotional schemes are employed in the market such that

the consumer's decision process is no longer solely a function of the product's price, then such a market fails to satisfy this condition, (ii) All consumers are indistinguishable from each other from the seller's point of view. If firm's have no other basis than the market price with which to decide who is to buy their product, then this requirement is also satisfied, (iii) The number of consumers and producers is sufficiently large so that the decisions of any one member of the market are not large enough to significantly alter the market price. If one buyer or seller is observed to be in a position to set his own price independently of the prevailing market price, then such an action would indicate that the market is not in a state of perfect competition, (iv) All consumers and producers are aware of current prices and bids for all the commodities in the market, (v) All consumers and producers are free to enter or leave the market as they see fit. There are no restrictions such as membership fees imposed upon this decision. If all these conditions are satisfied by a specific market under investigation then one would be entitled to apply the perfectly competitive market theory. Similarly, if these conditions were violated, but the market was such that it satisfied the requirements of another market type, i.e. monopoly, duopoly, oligopoly, oligopsony, etc., then the appropriate market theory could be employed.

However, each of these market theories is developed from the same basic deductive system. Hence, each theory is composed of propositions or hypotheses that require certain equilibrium conditions to be met as part of their initial conditions. Since the presence of equilibrium conditions within each theory is not dependent on the type of market under consideration, the market conditions become empirically relevant only if

the theories themselves contain hypotheses which can be subjected to test. But, the ubiquitous equilibrium conditions prevent these hypotheses from being submitted to test. Therefore, even though each market theory can only be empirically investigated if certain market conditions are satisfied, their inherent untestability is a consequence of the deductive system from which they are developed. Accordingly, the analysis and conclusions of this and the previous chapter apply to all theories developed in this manner. Moreover, until such time as testable theories of market behavior are constructed there is no need to investigate the empirical validity of the criteria which delimit particular market types.

4. Equilibrium Analysis and Economic Theory

The fundamental problem facing economists is to acquire a body of empirical knowledge about economic phenomena. Once this knowledge is acquired it can be used for whatever purposes economists or other social scientists have in mind. How such knowledge should be employed is not in question here. The primary object of the analysis has been the acquisitive process itself. In particular, the theories of classical economics have been examined to determine whether they permit the development of a set of testable theories of economic behavior. Without such theories there is no basis from which to develop empirical knowledge about economic events.

In the preceding sections economic theories of market behavior as well as their basic deductive system have been submitted to an extensive examination. The object has been to discover whether these theories can be submitted to test. The analysis has led to the conclusion that the classical

deductive system is such that it leads to the construction of theories which cannot be subjected to a process of refutation by empirical test. If a theory cannot be tested, it cannot be employed as the basis for acquiring empirical knowledge. Hence, we, as economists, are apparently in a position where it is not possible to employ classical economic theory as a basis from which to generate empirical knowledge of economic phenomena.

This is a strong as well as unfortunate conclusion. It implies that classical economic theory is empirically vacuous. It also states that classical theory cannot be employed to generate empirically significant explanations and predictions of economic events. Clearly, however, this theory was intended to provide economists with the ability to make empirically meaningful assertions about the relations among certain economic variables. Unfortunately, these intentions have not been translated by the method of equilibrium analysis into a body of testable theory. Since the primary objective remains unchanged--namely, to develop testable theories of economic behavior--it is necessary to delimit the conditions which if satisfied would provide classical economic theory with the requisite empirical content. To describe these requirements one needs to return for a moment to a general statement of what it is that one expects from a testable theory of economics.

In the second chapter the characteristics of such an economic theory are described by a statement of the classes of observable behavior that one expects it to encompass. H is the class of observed sequences of economic behavior whether in the past or the present that belong within the theory's domain. I is the class of all such observed sequences which includes those in H as well as all which will ever be professionally observed. If a theory

is testable within H then it is also testable within I. Whether it is empirically true or false can only be determined by actual tests. But, if theories are to have empirical significance then they must, in principle, have the ability to generate all of the relevant, observable sequences of behavior contained in H or I.

Classical economics contains theories that are stated in terms of equilibrium relations. Hence, if the class of observable sequences of economic behavior only includes behavior that is at equilibrium, then these theories are compatible with the behavior included in H or I. Unfortunately, economic theory does not provide sufficient interpretive rules to permit either the identification of sequences of equilibrium behavior or the testing of the relations in which these sequences occur. As a result, it is not possible to empirically establish whether H or I solely consist of sequences of equilibrium behavior. If one can neither establish nor refute a claim it has no empirical force. Consequently, the only conclusion to be drawn is that while H or I may include observable sequences of behavior at equilibrium such sequences do not exhaust the entire collection of observable behavior contained in H or I.

Consider, for a moment, an example^{5/} of such a state of affairs. Suppose the classes H and I consist of the observable sequences of behavior of a liquid in a specific container. If the container is always at rest then we could readily develop a theory which would account for the behavior

^{5/} I am indebted for this example to H.A. Simon, "Theories of Decision-Making in Economics and Behavioral Science," American Economic Review, Vol. 49, June, 1959, pp. 253-283.

recorded in H or I. Indeed, the theory could consist of a hypothesis which related the equilibrium position of the liquid's center of gravity to the gravitational forces acting upon it. In particular, the hypothesis could state that the liquid would minimize the height of its center of gravity. Accompanied by a statement of the initial conditions--in this case, the internal structure of the container--the equilibrium point is uniquely determined. Consequently, the theory is able to generate the observations contained in H or I.

If the container is not at rest, and if H and I still include all the liquid's observable behavior then the theory is no longer capable of generating these sequences of behavior. Clearly, if the container never comes to rest then the theory cannot, without further elaboration, generate any of the observed sequences of the liquid's behavior. If the theory does not contain interpretive rules with which we can determine when the liquid is in equilibrium, then we will be unable to test any of the theory's conclusions.

To describe and explain the behavior of this liquid we would need to know the processes by which it adapted itself to such changes in its environment as are brought about by the movement of the container. Once these processes are known and stated in a testable form then we can employ them to generate the behavior recorded in H or I.

In a similar manner equilibrium analysis provides us with theories which describe the end state of an equilibrating process. If the economic world remained at rest, and if the theory contained interpretive rules which permitted the observation of equilibrium points, then we could employ

classical equilibrium analysis to generate the observable, economic equilibrium states. Unfortunately, none of these conditions are met either by the classical theory or by the data which represent the observables noted in H or I.

To be able to describe and explain observable economic behavior we need to know the processes by which economic units whether individuals, firms, industries, or markets respond to their changing environment. To acquire a knowledge of these processes implies the existence of a theory or collection of theories which contain testable hypotheses about them. Such hypotheses must contain observable initial conditions. And if they are to permit the occurrence of economic events to be explained some of these hypotheses must be able to withstand and survive empirical tests.

In defence of classical theory it might be argued, that if the processes which guide the behavior of economic units are discovered then the equilibrium positions will turn out to be particular stages in the total sequence of observable behavior. Accordingly, while classical theory might not be able to generate all of the behavior contained in H and I it would be able to account for some. To argue in this fashion, however, is to overlook the fact that classical theory does not provide the criteria with which equilibria can be recognized. If the occurrence of equilibria cannot be detected then one cannot argue that one stage of a particular process represents an equilibrium point. Moreover, until such criteria are established it does not even make empirical sense to argue that equilibria occur at all. Further, a knowledge of the processes which determine the behavior of economic units does not imply that one will be able to

discover the equilibrium points of classical economic analysis. On the contrary, while a knowledge of the relevant processes would permit the generation of all the statements in H and I, it might also lead to the conclusion that classical equilibria, if empirically recognizable, do not exist as members of the total sequence of observable economic behavior.

Summary

At the beginning of this chapter it is noted that in order to explain the occurrence of an event a theory must contain at least one hypothesis which has survived empirical test. It is an obvious corollary of this statement that to meet this criterion the theory must contain at least one testable hypothesis. Since it was shown in some detail that the theory of market equilibrium did not contain such an hypothesis, the query was raised whether any of the theories of classical economics could satisfy this corollary.

To contain a testable hypothesis a theory must be stated in such a manner that the initial conditions of at least one hypothesis refer to observables. If all hypotheses of a theory contain as part of their initial conditions some terms which do not refer to observables then none of them can be submitted to test.

Classical economic theory is developed from a single deductive system. This deductive system employs a process of equilibrium analysis as the basis for generating economic theory. An analysis of this deductive system demonstrated that for every theory developed in this manner each of its hypotheses contained at least one equilibrium constraint as part of its initial conditions. Since the theories themselves do not provide criteria

by which these conditions can be empirically established, the equilibrium constraints do not refer to observables. Moreover, as none of the hypotheses have observable initial conditions, none of them can be submitted to empirical test. Therefore, it follows that none of these theories can be employed to explain or predict the occurrence of economic events.

The primary obstacle to empirical interpretation lies in the concept of an equilibrium. If a system is at rest or reaches a position of rest within a short interval of time, the end state may usefully be described as an equilibrium position. However, for systems that do not meet these conditions, the notion of an end state has less and less relevance the longer the time interval between states of rest. In observable economic systems the classical theories do not provide criteria with which it is possible to ascertain the presence of a state of rest. As a result, the classical economic conception of equilibrium has no empirical meaning within the context of observable economic behavior. If the concept of equilibrium is empirically vacuous then all theories based upon this concept must be vacuous as well.

In order to be able to describe and explain observable sequences of economic behavior one needs to be able to discover the processes by which the individual economic units respond to their changing environment. Once testable theories of these processes exist then it will be possible to proceed with the task of acquiring empirical knowledge about economic phenomena. Until such testable hypotheses are discovered no progress can be made. For although the object is to acquire knowledge, a necessary condition for such knowledge is the presence of testable hypotheses. Since classical equilibrium theory contains no such hypotheses it is necessary to continue the search by inspecting the hypotheses

and deductive system of otherbodies of economic theory. In this regard the next part of the book is devoted to an examination of the hypotheses and theories which result from econometric analysis.

PART THREE

Foundations and Characteristics

of

Econometric Analysis

Chapter 7

FOUNDATIONS OF ECONOMETRIC ANALYSIS

The theory of econometrics is to a large extent the theory of how to measure certain types of economic relations. While the subject of measurement is not sufficient by itself to differentiate econometrics from other parts of economics, the special techniques econometricians employ serve as the theoretical basis for their empirical investigations. One part of the econometrician's task is to observe actual economic data and measure the interactions among specific economic variables. The measuring procedures are a direct application of the theory of econometrics. The specific economic variables and their hypothesized relations to one another are derived in part from econometrics and in part from classical economic theory. Although the source of these relations is an important part of any econometric investigation, their structure is a consequence of the theory itself and not of the specific application.

Consider, for example, the following relation which represents a specific demand function.

$$y = \alpha + \beta p + u_t \tag{7.1}$$

where y is the quantity demanded, p is the prevailing price, α and β are parameters, and u_t is a random variable. The econometrician's task consists of first specifying that the relation (7.1) represents a certain demand function--i.e., the first step is to express the economic hypothesis or relation in a particular mathematical form. The next step is to employ such data as are available to derive estimates of the values of the parameters α and β , and of the error term u_t . With these estimates the relation in (7.1)

can now be confronted by additional data to determine, by means of certain statistical criteria, its "goodness-of-fit." If the goodness-of-fit is satisfactory the relation is then used as a vehicle for generating limited predictions about the future course of the relevant variables.

From this example, it is clear that the major part of the econometrician's job is encompassed by the first three stages; specifying the relations, estimating their parameters, and testing the degree to which the hypothesized relations fit the data. While the process of prediction is certainly valuable and important, it cannot be employed if the first three stages have not been successfully completed. As a result, an examination of the foundations of econometrics is primarily an analysis of the deductive system employed in the processes of specification, estimation, and testing.

If econometrics is to provide economists with the ability to establish testable theories of economic behavior, then its deductive system must permit the statement of hypotheses which can be subjected to test. If testable hypotheses can be generated by this approach then they can be examined to determine if any of them can survive the appropriate tests. Once tested relations are established we are at the beginnings of a science of economics. For with empirically tested hypotheses we can establish explanations and predictions of the relevant economic events.

The analysis of classical economics led to the conclusion that testable hypotheses were not a product of such a deductive system. It is the purpose of the following pages to discover whether econometrics is in the same class as classical economics, or whether its deductive system permits the development of a testable body of economic hypotheses. In order to clarify some of

the basic concepts of this deductive system the analysis begins with an examination of its fundamental assumptions or postulates.

1. The Basic Postulates

The first assumption of econometrics is that all hypotheses are to be stated in stochastic form. In the example above an error term u is included in the demand relation. This error term represents a random variable which is generated from a specific distribution with a known mean. The value of this error term varies from one interval of time to the next. Hence, a stochastic relation like (7.1) no longer states that the demand y is exactly equal to a constant α plus a constant β times price p , as in the case of classical theory. With the addition of the error term the deterministic relation is changed into an inexact specification of demand.

If one merely wished to construct an inexact relation there are many ways in which this could be accomplished. One could argue that the relation, $y = \alpha + \beta p$, is only a rough approximation to the actual relation governing consumer demand. At the same time one could claim that no matter how carefully data were collected these data would always contain errors.^{1/} Similarly, one could assert that no matter how many equations and variables were employed and despite the care with which the parameter values were estimated, the theories would never be better than inexact statements of hypothesized economic

^{1/}An excellent analysis of errors in economic data is to be found in: O. Morgenstern, On the Accuracy of Economic Observations, Second Edition, revised, Princeton University Press, Princeton, 1963.

relations. As a result, the equations could have errors in specification (an incomplete set of variables are selected), errors in measurement, or be subject to a variety of unspecified disturbances. To reduce this complexity of errors to manageable proportions, econometric theory assumes that measurement can be accomplished without error. If there are no errors in measurement, then a mis-specification of an hypothesis can be determined by testing its goodness-of-fit. Thus, all errors are categorized under the heading of random disturbances and are collected in the error term, u . Since the variables themselves are no longer supposed to be subject to variation, the relations are made stochastic by the addition of this error term.

There are many reasons for the addition of an error term into the structure of all econometric hypotheses^{2/}. But perhaps the fundamental justification lies in the statistical methods employed to estimate parameter values. If an equation contains a random variable as an error term--i.e. if we are dealing with stochastic relations--and if this random variable has certain well defined properties, then these properties delimit a particular class of estimating procedures. Since the process of parameter estimation depends upon the stochastic assumptions, different assumptions about the behavior of u entail different estimation techniques. The point to note, however, is that estimation procedures depend upon the stochastic assumptions. And since econometricians employ a specific set of estimation techniques these procedures imply that econometric hypotheses must contain error terms with certain specific characteristics.

^{2/}A list of such reasons is provided in: S. Valavanis, Econometrics, McGraw-Hill, New York, 1959, pp. 5-6.

A. The Error Term

The characteristics of an error term are a consequence of the assumptions that are made about its behavior. While a number of different assumptions can be and are employed in a variety of econometric studies the set to be examined are those which are most generally used.

The first assumption is that for every value of t u_t is a random variable. A random variable can be discrete or continuous, but in each case it must take on each of its several values with a definite probability. In mathematical terms this assumption is stated as follows:

If $p(u)$ is defined for all values of u such that

$$\int_u p(u) du = 1$$

or

$$\sum_u p(u) = 1$$

where $0 \leq p \leq 1$

then u_t is a random variable for all values of t .

In equation (7.1) this assumption states that at each period of time the value of u_t is determined by its own density function. In other words, while the value of u_t may change from period to period the actual value is determined by its probability of occurrence which is defined by its density or probability function. If this density function is unknown, then the probabilities with which u_t takes on specific values are also unknown. Further, it is not possible to analytically determine any of its characteristics, i.e. its expected value, variance, and higher moments. Hence, if one is to be able to specify the properties of u_t its density function as well as the values of certain of its moments must be known.

The second assumption concerns the density function and states that $\underline{u}_t$ is normally distributed. Mathematically this assumption is expressed by

$$p(u) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{u_t - E[u]}{\sigma} \right)^2} \quad (7.2)$$

where $-\infty < u < \infty$

and where $E[u]$ represents the mean and σ the standard deviation of the density function. Having specified the density function one now needs to know what values to associate with its mean and variance.

The third assumption is that the expected value of $\underline{u}_t$ is equal to zero--i.e.

$$E[u_t] = \int_{-\infty}^{\infty} u_t p(u) du = 0 \quad (7.3)$$

for all t .

Combining (7.3) with (7.2) the density function of $\underline{u}_t$ can be rewritten as

$$p(u) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{u_t}{\sigma} \right)^2} \quad (7.4)$$

To be able to estimate the value of $\underline{u}_t$ at any particular period of time and to relate this value to one derived during another time interval, it is necessary to know that the variance of $\underline{u}_t$'s density function is not changing over time. Hence, the fourth assumption states that $p(u)$ has a finite variance which is constant over time, i.e. the second moment about the mean, σ_t^2 , is constant over time and finite:

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 \quad (t=1,2,3,\dots)$$

and $0 < \sigma_t^2 < \infty$

where $\sigma_t^2 = \int_{-\infty}^{\infty} (u_t - E[u_t])^2 p(u) du$

One consequence of the last three assumptions is that the value of the error term for any particular period is derived by a random selection from a normal distribution with zero mean and constant variance. Each value of $\underline{u}_t$ does not depend on any of its preceding or succeeding values. Each is independent of each other and depends solely on the particular normal distribution from which it is derived. As a result, there are two further assumptions implicit in this characterization of the error term.

The fifth assumption is, as has just been noted, that specific values of $\underline{u}_t$ are independent of each other, i.e. are not correlated to one another. Mathematically this statement is represented by the condition:

$$E[u_t, u_{t-i}] = 0$$

where $\underline{t}$ may take on all possible values and $i \neq 0$

Concurrently, if all values of $\underline{u}_t$ are to be independent of each other, they must also be independent variable contained in the relevant hypotheses. For relation (7.1) this last assumption states that the covariance of $\underline{u}_t$ and $\underline{p}$ must be equal to zero, i.e. $\underline{u}_t$ is independent of $\underline{p}$ if

$$\text{Cov}(u_t, p_{t-i}) = \sigma_{u_t, p_{t-i}} = 0$$

for all $\underline{t}$ and all $\underline{i}$

These assumptions delimiting the properties of the error term are stated for the case where a theory consists of one equation. But, not all theories are quite so simple and econometricians frequently deal with theories containing several equations. Since each relation contains an error term these assumptions need to be interpreted to include this general case.

Suppose for the moment that we are dealing with the general case where there are $\underline{n}$ relations containing $\underline{n}$ variables. Since each relation is stochastic,

each relation has an error term. For the $\underline{n}$ relations there will be $\underline{n}$ such error terms which can be represented by a vector $\underline{U}(t) = (u_1(t), u_2(t), \dots, u_n(t))$.

If $n=1$, the first assumption requires that $u(t)$ be a random variable. If $n \neq 1$, all $u_j(t)$, ($j=1, 2, \dots, n$) are random variables. Hence, the vector $\underline{U}(t)$ is also a random variable.

The second assumption is generalized by requiring the error terms $(u_1(t), u_2(t), \dots, u_n(t))$ to be jointly normally distributed.

Similarly if for $n=1$, $E[u(t)] = 0$, then the means of all the error terms are equal to zero. Consequently, the vector given by $E[U(t)]$ is a vector of zeros, i.e. $E[U(t)] = (0, 0, \dots, 0)$. This has the same meaning as the statement that the joint normal distribution has a mean equal to zero.

If the joint normal distribution is to have a constant variance then the covariances of the respective errors must not vary with time, i.e.

$$\text{Cov}(u_j(t), u_k(t)) = \sigma_{jk}(t) = \text{constant for all } t$$

If each of the values of the $\underline{n}$ error terms is to be independent of all of its other values then for each relation

$$E[u_j(t), u_j(t-i)] = 0$$

$$\text{for all values of } t \text{ where } i \neq 0 \text{ and } j=(1, 2, \dots, n)$$

Similarly, if the vector $\underline{U}(t)$ is to be independent of the values of any of the variables in the $\underline{n}$ relations, the covariances between each error term and the independent variables in its equation must also be equal to zero.

Earlier it is noted that estimation procedures depend upon the stochastic assumptions. Having described the assumptions concerning the behavior of the error term it is now sensible to examine the statistical procedures employed in the estimation of the parameters of these stochastic relations.

2. Estimating the Parameters

A. Single Equation Models

In order to estimate the parameters of a particular hypotheses it is necessary to have an estimating criterion function. While a number of such functions can be generated and used, the principal criterion employed in econometrics is the maximum likelihood criterion.

All estimating functions produce estimates which have somewhat different properties. But very few are able to generate estimates that are unbiased, consistent, sufficient, and efficient^{3/}. In most normal situations the maximum likelihood criterion produces estimates with these characteristics. As a result of these as well as other convenient properties the majority of estimates in econometrics are maximum likelihood estimates.

To explicate the notion of a maximum of likelihood consider a simple estimation problem^{4/}. Suppose we are faced with an urn in which there are a number of red and white balls. Suppose further that we know there are twice as many of one color as of the other, but we do not know which color is the more numerous. If we draw a sample of n balls from the urn with replacement we know that the distribution of the number of white balls in the urn is given by the binomial

$$f(x; p) = \binom{n}{x} p^x q^{n-x}$$

Further, we know that the probability of drawing a white ball is either 1/3 or

^{3/}A definition of these terms is given in the mathematical Appendix B.

^{4/}This example is adapted from: A.M. Mood and F.A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill, New York, Second Edition, 1963, p. 179.

2/3. The problem is to estimate the value of p from the particular colors of the balls in a specific sample. If the sample consists of four balls, then the total number of possible outcomes is given by:

no. of white (x)	0	1	2	3	4
$f(x, 1/3)$	16/81	32/81	24/81	8/81	1/81
$f(x, 2/3)$	1/81	8/81	24/81	32/81	16/81

Consequently, if only one of the balls in the sample is white the estimate of $p=1/3$ would be chosen, since the probability of 32/81 is greater than that of 8/81. Similarly if the sample contained three white balls our estimate of p would be 2/3.

For the procedure to provide a good estimate of p it is clearly necessary that the sample represent the distribution of balls in the urn. If we draw a sample with no white balls when there are twice as many white as red, the sample would lead to an erroneous estimate of p . The maximum likelihood procedure assumes that the sample is representative of the population. Hence, in this example, for every value of x it selects the value of p so that

$$f(x; \hat{p}) > f(x; p')$$

where $\hat{p}$ represents the maximum likelihood estimate, and p' the alternative value. To generalize this result to the case where an estimate is required of the value of an unknown parameter u from a random sample of size n the procedure is as follows. The sample values $x_1, x_2, \dots, x_n$ define a sample density $f(x_1, x_2, \dots, x_n; u)$. The maximum likelihood estimate of u is the number $\hat{u}$, if it exists, such that the value of

$$f(x_1, x_2, \dots, x_n; \hat{u}) > f(x_1, x_2, \dots, x_n; u')$$

where u' is any other possible value of u .

In order to find a particular maximum likelihood estimate it is necessary to construct the likelihood function and then find its maximum point. The likelihood function is derived from the sample density. A sample value from a population of known density provides us with the sample density for that value. For example, two white balls in a sample of five drawn from the urn gives a sample density of $f(2; p) = \binom{5}{2} p^2 (1-p)^3$, where $f(2; p)$ is the likelihood function in this case. If p could take on all possible values, i.e. $0 \leq p \leq 1$, then to find the value of p which maximizes the likelihood function one would differentiate $f(2; p)$ with respect to p , set the resulting equation equal to zero, and solve for p . The solution is $\hat{p}$, the maximum likelihood estimate of p .

In general, if $x_1, x_2, \dots, x_n$ are sample values and $f(x_1, x_2, \dots, x_n, u)$ is the sample density, then the function $\prod_{i=1}^n f(x_i; u)$ is the likelihood function of u for the particular sample values $(x_1, x_2, \dots, x_n)$. Since $\prod_{i=1}^n f(x_i; u)$ has its maximum at the same point as the function, $\log \prod_{i=1}^n f(x_i; u)$, and because the logarithm of the likelihood function is usually easier to deal with, it is customary to find the maximum of the logarithm of the likelihood function.^{5/}

To apply this procedure to an econometric relation consider the demand relation, $y_t = \alpha + \beta p_t + u_t$. By a suitable transposition of terms this equation becomes

$$u_t = y_t - \alpha - \beta p_t \quad (7.5)$$

According to the sixth assumption about the error term, u_t must not be directly dependent upon any of the independent (exogenous) variables. From (7.5) it is clear that, while $\text{Cov}(u_t, p_{t-1})$ may be equal to zero, the value of u_t

^{5/} For a more detailed presentation of the maximum likelihood technique see: A.M. Mood and F.A. Graybill, op. cit., Chapter 8.

is a function of the observed values of $\underline{y}_t$ and $\underline{p}_t$ as well as of the unknown values of $\underline{\alpha}$ and $\underline{\beta}$.

The problem is to derive maximum likelihood estimates of $\underline{\alpha}$ and $\underline{\beta}$. It is assumed, as in the urn example, that the sample data are representative of the population. Hence, one begins by observing a sample of values of $\underline{y}$ and $\underline{p}$ at a particular time $\underline{t}$. To construct the likelihood function one needs to know the sample distribution of $\underline{u}$. But the assumption about the error terms is that the $\underline{u}$'s are normally distributed with zero mean and constant variance. As the sample is assumed to be representative of the population, the sample distribution of $\underline{u}$ is the normal distribution with zero mean and fixed variance. Consequently, the likelihood function is a function of the values of $\underline{u}$ which correspond to the observed sample values of $\underline{y}$ and $\underline{p}$. Denoting the likelihood function by the letter $\underline{L}$ a sample of size $\underline{n}$ gives the likelihood function:

$$\underline{L} = f(u_1, u_2, \dots, u_n)$$

where the particular values of $\underline{u}_i$ ($i=1,2,\dots,n$) are a function of the $\underline{n}$ pairs of observations on $\underline{y}$ and $\underline{p}$. To maximize $\underline{L}$ one has to find those values of $\underline{\alpha}$ and $\underline{\beta}$ which make $\underline{L}$ as large as possible.

Each sample value of $\underline{u}$ is normally distributed. Thus, the distribution of $\underline{n}$ sample values is a multivariate normal. The multivariate normal relates to the univariate normal distribution in the following way. The univariate normal is customarily written as in (7.2):

$$p(u) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{u - E[u]}{\sigma} \right)^2}$$

However, in order to point out the relation between the univariate and the multivariate normal, the univariate distribution can be stated as

$$p(u) = \frac{1}{\sqrt{2\pi} \sigma} \frac{1}{\sqrt{|\sigma_{u_i u_j}|}} e^{-\frac{1}{2}\{(u-E[u])(\sigma_{u_i u_j})^{-1}(u-E[u])\}} \quad (7.6)$$

where $|\sigma_{u_i u_j}|$ represents the determinant of covariances of u_i and u_j , and $(\sigma_{u_i u_j})^{-1}$ represents the inverse of the covariance of u_i and u_j . In the case of one variable the value of $|\sigma_{u_i u_j}|$ is the same as the value of σ_u . For the determinant only contains this element. Similarly for the univariate normal $(\sigma_{u_i u_j})^{-1}$ is equal to $(\sigma_u)^{-1}$.

To derive the multivariate normal one must designate the n variables. Let $u_1, u_2, \dots, u_n$ be n variables which have a joint normal distribution. Then, as noted above, these variables can be represented by the vector $\underline{U} = (u_1, u_2, \dots, u_n)$. In a similar fashion the expected values of these variables can be represented by the vector $\underline{E[U]} = (E[u_1], E[u_2], \dots, E[u_n])$. For n variables the matrix of covariances is given by:

$$\sigma_{UU} = \begin{bmatrix} \sigma_{u_1 u_1} & \sigma_{u_1 u_2} & \dots & \sigma_{u_1 u_n} \\ \sigma_{u_2 u_1} & \sigma_{u_2 u_2} & \dots & \sigma_{u_2 u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_n u_1} & \sigma_{u_n u_2} & \dots & \sigma_{u_n u_n} \end{bmatrix} \quad (7.7)$$

Hence the value of $|\sigma_{UU}|$ is the determinantal value of the matrix given in (7.7), and the value of $(\sigma_{UU})^{-1}$ is the value of the inverse of (7.7).

Employing this notation the multivariate normal is given by

$$p(U) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sqrt{|\sigma_{UU}|}}\right) e^{-\frac{1}{2}\{(U-E[U])(\sigma_{UU})^{-1}(U-E[U])\}} \quad (7.8)$$

Returning to the estimation problem one can now write the likelihood function for a sample of n values of the u 's as.

$$L = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\sqrt{|\sigma_{UU}|}}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\{(U-E[U])(\sigma_{UU})^{-1}(U-E[U])\}} \quad (7.9)$$

However, the third assumption about the error term states that the expected value of each u is zero, i.e. $E[U] = (0,0,\dots,0)$. Therefore, (7.9) becomes

$$L = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{|\sigma_{UU}|}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\{U(\sigma_{UU})^{-1}U\}} \quad (7.10)$$

Further the fifth assumption states that the covariances between each error term and each independent variable is zero, i.e. $\sigma_{u_i u_j} = 0$ for $i \neq j$. Consequently, the matrix σ_{UU} is given by

$$\sigma_{UU} = \begin{bmatrix} \sigma_{u_1 u_1} & 0 & \dots & 0 \\ 0 & \sigma_{u_2 u_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{u_n u_n} \end{bmatrix} \quad (7.11)$$

Thus, the value of the determinant $|\sigma_{UU}|$ is the product $(\sigma_{u_1 u_1})(\sigma_{u_2 u_2})\dots(\sigma_{u_n u_n})$. But, by the fourth assumption $\sigma_{u_1 u_1} = \sigma_{u_2 u_2} = \dots = \sigma_{u_n u_n}$. Hence, the value of $|\sigma_{UU}| = (\sigma_{uu})^n$ or $(\sigma_u)^n$. Similarly, the value of the inverse $(\sigma_{UU})^{-1}$ is given by the inverse $(\sigma_{uu})^{-1}$ or $(\sigma_u)^{-1}$. Accordingly, (7.10) can be simplified to

$$L = (2\pi)^{-\frac{n}{2}} (\sigma_u)^{-\frac{n}{2}} e^{-\frac{1}{2}\{U(\sigma_u)^{-1}U\}} \quad (7.12)$$

Taking the logarithm of the likelihood function (7.12) the function becomes

$$\log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_u - \frac{1}{2} (\sigma_u)^{-1} \sum_{i=1}^n u_i^2 \quad (7.13)$$

To find the maximum of a function one takes the partial derivatives with respect to the unknown variables, sets the resulting equations equal to zero, and

solves for the unknown values.^{6/} In equation (7.13) the unknowns are σ_u and the u_i . But, $u_i = y_i - \alpha - \beta p_i$. Substituting this value for u_i into (7.13) the logarithm of the likelihood function is given by:

$$\log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_u - \frac{1}{2} (\sigma_u)^{-1} \sum_{i=1}^n (y_i - \alpha - \beta p_i)^2 \quad (7.14)$$

The unknowns are now α , β , and σ_u , since y_i and p_i are the observed sample values. By taking partial differentials of (7.14) with respect to α , β , and σ_u and by setting the resulting equations equal to zero one generates three equations in three unknowns

$$\begin{aligned} \sum_{i=1}^n (y_i - \alpha - \beta p_i) &= 0 \\ \sum_{i=1}^n (y_i - \alpha - \beta p_i) p_i &= 0 \\ \frac{1}{n} \sum_{i=1}^n (y_i - \alpha - \beta p_i)^2 &= \sigma_u \end{aligned} \quad (7.15)$$

Solving the equations in (7.15) in terms of the three unknowns α , β , and σ_u produces the maximum likelihood estimates which are designated by $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}_u$.

While the equation used in this example contains only one independent variable, p_t , the procedure for deriving maximum likelihood estimates is the same for equations with n independent variables. For example, if the demand for a particular commodity can be represented as a function of the prices of n other commodities, then this hypothesis is represented by:

$$y = \alpha + \beta_1 p_1 + \beta_2 p_2 + \dots + \beta_n p_n + u_t$$

The error term is still a function of these $n+1$ variables, i.e. $u = y - \alpha - \beta_1 p_1 - \beta_2 p_2 - \dots - \beta_n p_n$ and the covariance of u_t with each of these variables must be equal to zero.

^{6/} Likelihood functions do not have a minimum, hence it is not necessary to apply the second order condition for a maximum.

Accordingly, the likelihood function is formed in the manner described above and the partial differential equations are solved in the standard fashion for the maximum likelihood estimates $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n, \hat{\sigma}_u$.

It should be noted that when the parameter values of a single equation are estimated the maximum likelihood estimates are equivalent to the estimates derived by the method of least squares. That is to say, the equations in (7.15) are identical to the estimating equations generated by the least squares approach. However, as soon as the parameters of more than one equation are being simultaneously estimated the equivalence between least squares and maximum likelihood estimates no longer holds.

B. Several Equations Models

In order to examine the case where the econometric theory consists of several equations, consider the general model which contains the following n equations:

$$\begin{aligned} y_1 + \alpha_{12}y_2 + \dots + \alpha_{1k}y_k + \beta_{11}z_1 + \beta_{12}z_2 + \dots + \beta_{1m}z_m &= u_1 \\ \alpha_{21}y_1 + y_2 + \dots + \alpha_{2k}y_k + \beta_{21}z_1 + \beta_{22}z_2 + \dots + \beta_{2m}z_m &= u_2 \\ \vdots & \\ \alpha_{k1}y_1 + \alpha_{k2}y_2 + \dots + y_k + \beta_{k1}z_1 + \beta_{k2}z_2 + \dots + \beta_{km}z_m &= u_k \end{aligned} \quad (7.16)$$

In (7.16) the y 's represent the dependent (endogenous) variables, the z 's the independent (exogenous) variables, the α 's the coefficients of the endogenous variables, the β 's the coefficients of the exogenous variables, and the u 's are the error terms. To estimate the α 's and β 's by the maximum likelihood approach it is necessary to be sure that the error terms of one equation are not correlated with the error term of any other equation,---i.e. the $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$.

Earlier, when the case of several observations on one equation was considered it was noted that the covariance between the several values of the one error term had to be equal to zero. With more than one equation there are a number of different error terms. As a result, in order to ensure that the value of each error term is independent of the value of any other error term, the covariance between any two error terms must also be equal to zero.

Two further points need to be mentioned about the system of equations in (7.16). The first is that none of these equations may represent an economic or accounting identity. One obvious example of an economic identity is the equation

$$c(t) + s(t) = y(t) \quad (7.17)$$

where $y(t)$ represents income, $c(t)$ consumption, and $s(t)$ savings. While the model (system of equations) might contain a function relating consumption to income, e.g. $c(t) = \alpha + \beta y(t) + v(t)$, the presence of the strict identity violates the requirement that the value of $u(t)$ be uncorrelated with the values of the independent variable $y(t)$. For α and β are constants and $c(t)$ and $s(t)$ are the dependent variables. Hence, under this arrangement $u(t)$ determines the value of $y(t)$. Accordingly, the presence of identities in a system such as (7.16) violates the requirement that the covariance between the error term and the independent variables must be zero.

The second point to note in (7.16) is that the equations are written without a constant term and with parameters $\alpha_{ij} = 1$ for $i=j$. The constant term can be included by letting the final independent variable, z_{im} , be equal to 1. In this case its parameter β_{im} represents the constant term. Concurrently, it is possible to reduce the number of parameters by dividing each equation by one of its coefficients--namely, α_{ij} , where $i=j$.

To form the likelihood function of this system of equations it is convenient to make use of some simplifying notation. If the endogenous variables are separated from the exogenous in equation (7.16) the coefficients of these variables can be represented by two matrices:

$$A = \begin{bmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & 1 & \cdots & \alpha_{2k} \\ \vdots & \vdots & & \vdots \\ \alpha_{k1} & \alpha_{k2} & \cdots & 1 \end{bmatrix} \quad B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{km} \end{bmatrix}$$

Further, if the matrix of endogenous variables is represented by $\underline{Y}$ and the corresponding matrix of exogenous variables by $\underline{Z}$, then the equations in (7.16) can be written in the simplified matrix form

$$AY + BZ = U \quad (7.18)$$

where U is the vector of error terms $\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}$.

In order to estimate the values of the coefficients in A and B a sample of observations is needed for each of the rows of the matrices $\underline{Y}$ and $\underline{Z}$. Suppose that the samples are all of the same size with each containing $\underline{n}$ observations. The likelihood function of the sample can then be written as:

$$L = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} |J|^N \left(\frac{1}{|\delta_{km}|}\right)^{\frac{N}{2}} e^{-\frac{1}{2} \left\{ \sum_{n=1}^N U_n [\delta_{km}]^{-1} U_n' \right\}} \quad (7.19)$$

Except for the addition of the term $|J|^N$, this relation has exactly the same form as the likelihood function in (7.10). In (7.10) the term, $\underline{\sigma}_{UU}$, represents the diagonal matrix of covariances given in (7.11). In (7.19) the term, $\underline{\delta}_{km}$, represents the matrix of covariances given by $\underline{\text{Cov}}(u_i, u_j)$. But as is noted above $\underline{\text{Cov}}(u_i, u_j) = 0$

for all $i \neq j$. Therefore, $\underline{\delta}_{km}$, is also a diagonal matrix of covariances, with $|\delta_{km}|$ representing its determinantal value, and $[\delta_{km}]^{-1}$ its inverse.

The term $|\underline{J}|$ is the determinant of the Jacobian matrix.^{7/} When one is dealing with a single linear equation the value of $|\underline{J}|$ is 1, and it can be ignored as in (7.10). However, when the model contains two or more equations the matrix of these partial derivatives must be included.

To evaluate $|\underline{J}|$ two conditions must be met. The first is that there must be a one-to-one functional relation between the error terms and the dependent variables. If the relation is one-to-many then the Jacobian is undefined. Since $\underline{J}$ is a matrix of partial differentials, e.g. $\partial u / \partial y$, the functions must also be continuous with first derivatives defined over the relevant domain. For $|\underline{J}|$ to have a unique value $\underline{J}$ must be non-singular. Hence, the second condition is that $|\underline{J}|$ can only be evaluated if there are as many u 's as there are y 's.

If these conditions are met and if the model can be represented by a set of linear equations as in (7.16) then the matrix of partial derivatives given by $\underline{J}$ is the matrix of coefficients of the y 's, i.e.

$$\underline{J} = \begin{bmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & 1 & \cdots & \alpha_{2k} \\ \vdots & \vdots & & \vdots \\ \alpha_{k1} & \alpha_{k2} & \cdots & 1 \end{bmatrix} = \underline{A}$$

Accordingly, the likelihood function for the equations in (7.16) is given by:

$$L = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} |\underline{A}|^N \left(\frac{1}{|\delta_{km}|}\right)^{\frac{N}{2}} e^{-\frac{1}{2} \left\{ \sum_{n=1}^N u_n [\delta_{km}]^{-1} u_n' \right\}} \quad (7.20)$$

It follows, then that the logarithm of the likelihood function can be represented by:

^{7/} Jacobians are discussed in greater detail in: L.R. Klein, Textbook of Econometrics, Row, Peterson and Company, Evanston, 1953, pp. 32-38.

$$\log L = -\frac{N}{2} \log 2\pi + N \log |A| - \frac{N}{2} \log |\delta_{km}| - \frac{1}{2} \sum_{n=1}^N U_n [\delta_{km}]^{-1} U_n' \quad (7.21)$$

Now (7.18) states that, $U = AY + BZ$. Thus, to find the maximum likelihood estimates of the $\underline{\alpha}$'s, $\underline{\beta}$'s, and the covariances (u_i, u_j) for $i=j$, one substitutes, $U = AY + BZ$, into (7.21) to get the likelihood function in the same form as (7.14), i.e.

$$\begin{aligned} \log L = & -\frac{N}{2} \log 2\pi + N \log |A| - \frac{N}{2} \log |\delta_{km}| \\ & - \frac{1}{2} \sum_{n=1}^N (AY+BZ)_n [\delta_{km}]^{-1} (AY+BZ)_n' \end{aligned} \quad (7.22)$$

To derive the system of estimating equations analogous to (7.15) one takes the partial differentials of (7.22) with respect to the $\underline{\alpha}$'s, $\underline{\beta}$'s, and $\underline{\sigma}_{uu}$'s and sets the resulting relations equal to zero. The simultaneous solution of these equations generates the maximum likelihood estimates, $\hat{\underline{\alpha}}$'s, $\hat{\underline{\beta}}$'s, and $\hat{\underline{\sigma}}$'s.

Since the $\hat{\underline{\alpha}}$'s, $\hat{\underline{\beta}}$'s, and $\hat{\underline{\sigma}}$'s are a consequence of the simultaneous solution of the estimating equations each estimate, say $\hat{\alpha}_{12}$, depends on the values of the remaining parameters. That is to say, the estimates of the $\hat{\underline{\alpha}}$'s for one equation depend on the values for the $\hat{\underline{\alpha}}$'s and $\hat{\underline{\beta}}$'s of another equation; and the estimates of the $\hat{\underline{\sigma}}$'s depend on one another as well as the estimates of the $\hat{\underline{\alpha}}$'s and $\hat{\underline{\beta}}$'s of another equation.

When dealing with a single equation model it was noted that the method of least squares produced exactly the same estimating equations as the maximum likelihood approach. In this situation, both methods are equally appropriate for estimating the parameters of a single equation. However, the moment models are entertained which contain several equations the standard least squares approach can no longer be employed. The reason for this is quite simple. If the method of least squares is employed to estimate the parameters in (7.16) each equation

is treated as though it were independent of the remainder. Such a procedure ensures that the values of the parameters of one equation do not affect those of another.

For example, by using the least squares approach the covariances σ_{uu} of each equation would be estimated without reference either to the remaining covariances or to the values of the $\hat{\alpha}$'s and $\hat{\beta}$'s of the other equations. Similarly, the term $|J|=|A|$ would be excluded from the estimating equations. As a result, by applying the method of least squares to a model containing several equations estimates would be generated that are completely at variance with the maximum likelihood estimates. Since the maximum likelihood criterion usually provides estimates which have the desirable characteristics of unbiasedness, consistency, sufficiency and efficiency, least squares is clearly inappropriate except in the case where the parameters of a single equation are being estimated.

The correct procedure is to employ the function given in (7.22) to generate the estimation equations. This process, however, involves a large amount of computation. To reduce the computations various special techniques are employed. Because the method of least squares is not a viable alternative, a number of approximations to maximum likelihood have evolved. Among these the techniques of limited information, instrumental variables, and Theil's method of reduced forms are frequently used.^{8/} The object of these methods is to approximate the

^{8/}For a detailed discussion of these techniques see: S. Valavanis, op.cit., Chapters 7, 8, and 9.

maximum likelihood estimates without going through the elaborate analysis required by the solution of the equations derived from (7.22). While such techniques are important to the practicing econometrician they can be disregarded here as this analysis is concerned with the foundations of measurement in econometrics not with various computational procedures.

3. Testing Statistical Hypotheses^{9/}

A statistical hypothesis is a statement about the probability density function (frequency function) of a random variable. Since all econometric hypotheses contain a random variable with an assumed density function, all such statements are statistical hypotheses. In order to submit such hypotheses to empirical test a procedure is required which permits the decision to be made whether to accept or reject a particular hypothesis.

Suppose for the moment that one is interested in the specific hypothesis

$$y(t) = \alpha + \beta p(t) + u(t) \quad (7.23)$$

where $u(t)$ is normally distributed with zero mean and constant variance. To estimate the values of α and β one takes a sample of observations on $y(t)$ and $p(t)$ and proceeds, in this case by either least squares or maximum likelihood, to generate the estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}$. These estimates are based on a single, specific set of observations. Hence, without the introduction of additional criteria, it is not possible to assess the reliability of these estimates. For example, assume in the context of a particular set of data that the values of the estimates are given by: $\hat{\alpha}=1.02$, $\hat{\beta}=2.31$ and $\hat{\sigma}=3.40$. If these values are substituted directly

^{9/}An extensive discussion of the problems surrounding statistical testing is to be found in: C.W. Churchman, Theory of Experimental Inference, Macmillan, New York, 1948.

into (7.23) a specific application of the hypothesis to a particular set of data is produced. But without estimating the reliability of these estimated values it is not possible to determine whether to accept or reject (7.23) as a statement of the functional relation governing y(t) and p(t).

In order to assess the statistical reliability of an estimate one needs to examine its possible range of variation. Parameter estimates are based on finite samples. Also each sample generates a different value for the parameter being estimated. As a result, each estimate contains a certain error which is ascribed to the sampling process. If the sample was infinite and included all possible observations on the relevant variables, the estimated value would equal the true parameter value. But, all econometric samples are finite. Thus, the first task is to estimate the size of the sampling error.

One such measure is provided by computing the variance of the parameter estimate. If $\hat{\beta}$ is the maximum likelihood estimate of β , then $\hat{\beta}$ is a function of the sample observations and is itself a random variable. The variance of a random variable is the expected value of the square of the difference between the random variable and its expected value. Accordingly, the variance of $\hat{\beta}$ is given by:

$$\sigma_{\hat{\beta}}^2 = E[\hat{\beta} - E[\hat{\beta}]]^2 \quad (7.24)$$

But $\hat{\beta}$ is an unbiased estimate of β . That is to say, if a series of random samples were drawn and an estimate of β were computed from each a distribution of values of $\hat{\beta}$'s would result. Since $\hat{\beta}$ is unbiased the mean of this distribution of $\hat{\beta}$'s would approach the true value of β . In the limit as the number of samples increased to infinity, $E[\hat{\beta}] = \beta$, where the individual values of $\hat{\beta}$ are dispersed about this mean with a variance given by (7.24). Because $E[\hat{\beta}] = \beta$, (7.24) can be rewritten as:

$$\sigma_{\hat{\beta}}^2 = E[\hat{\beta} - \beta]^2 \quad (7.25)$$

when the sample estimates of β are normally distributed one can expect approximately 95% of any set of values, $\hat{\beta}_i$, ($i=1,2,\dots,n$) to lie between the two limits

$$\beta - 2\sigma_{\hat{\beta}} \quad \text{and} \quad \beta + 2\sigma_{\hat{\beta}} .$$

In econometrics it is usual only to have a single sample and as a result a single estimate of β . If the sample size is sufficiently large ($n > 30$), and $\hat{\beta}$ can be assumed to be normally distributed, then the 95% confidence interval for the estimate is provided by the limits, $\hat{\beta} - 2\sigma_{\hat{\beta}}$ and $\hat{\beta} + 2\sigma_{\hat{\beta}}$. If the sample size is smaller than 30 the limits are formed by using the t distribution e.g. $\hat{\beta} - t_{.05}\sigma_{\hat{\beta}}$ and $\hat{\beta} + t_{.05}\sigma_{\hat{\beta}}$, where the value of $t_{.05}$ is found from tables of the t distribution at this level of significance and the relevant sample size.^{10/} To construct and employ these confidence intervals $\hat{\beta}$ must be normally distributed. In (7.23) the distribution of the values of $\hat{\beta}_i$ is a function of the sample observations of y_i ; and the distribution of the β_i is a linear function of the values of y_i . In turn, y_i is a linear function of the values of the random variable u_i . And u_i is normally distributed with zero mean and constant variance. Hence both y_i and $\hat{\beta}_i$ are normally distributed and it is legitimate to employ the particular confidence intervals noted above.

Applying these results to the example at the beginning of this section it is now possible to estimate the reliability of the parameter estimates. All that is required is to compute the standard error of each estimate, i.e. $\sigma_{\hat{\alpha}}$, $\sigma_{\hat{\beta}}$, and $\sigma_{\hat{\theta}}$, and employ the appropriate confidence limits. For example, the standard

^{10/} The properties of these and other sample distributions are discussed in: A.M. Mood and F.A. Graybill, op. cit., Chapter 10.

error for, $\sigma_{\hat{\beta}}$, in (7.23) is given by the formula

$$\sigma_{\hat{\beta}} = \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (p_i - \bar{p})^2}} \quad (7.26)$$

where σ^2 is the variance of u , p_i the sample values of the exogenous variable p , and $\bar{p}$ the mean of these sample values. Now the variance of u is an unknown quantity which has a calculated estimate, $\hat{\sigma}^2$. Substituting $\hat{\sigma}^2$ for σ^2 in (7.26) one arrives at a formula which provides the sample estimate of the standard error.^{11/}

$$\sigma_{\hat{\beta}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (p_i - \bar{p})^2}} \quad (7.27)$$

From (7.27) one can directly compute the confidence limits which for $n > 30$ and a 95% level of confidence are $\hat{\beta} \pm 2\hat{\sigma}_{\hat{\beta}}$.

Having defined a measure of reliability it is now possible to return to the problem of describing a procedure by which one can decide whether to accept or reject a particular hypothesis. The theory of hypothesis testing is customarily phrased in terms of a choice between two alternative hypotheses: the null hypothesis, H_0 , and some alternative hypothesis H_1 . In the above example, the null hypothesis could be that the true value of $\beta = 1.5$, i.e.

$$H_0: \beta = 1.5$$

The alternative hypothesis could be that β has another specific value or just

$$H_1: \beta \neq 1.5$$

^{11/} This formula as well as the formula for the general case is derived in: L.R. Klein, op. cit., pp. 134-137.

To be able to decide whether to accept H_0 a critical or rejection region for the values of β must be defined. If the observed (estimated) value of β falls within this critical region H_0 is rejected. Otherwise it is accepted. The normal procedure is to set the acceptance region so that it includes 95% of the area of the function defined by H_0 . By doing so one incurs the chance of committing a Type I error 5% of the time. That is to say, 5% of the time it is possible to reject H_0 when it is in fact true. Similarly, there is a certain chance, the Type II error, that H_0 will be accepted when it is false. One cannot avoid making these errors. Consequently, to select the best of a set of alternative tests, one chooses that test which for a specified Type I error has the smallest Type II error.^{12/}

In order to clarify this procedure consider once again the example mentioned at the beginning of this section. Suppose that a sample of forty observations are taken and that the estimate for $\hat{\beta}$ is given by $\hat{\beta} = 2.31$ (see above). The null hypothesis is $H_0: \beta = 1.5$. The Type I error is set at 5%, i.e. (a 95% level of confidence is adopted) and the test can now be stated as follows: If the value of $\beta = 1.5$ is contained in the interval $\hat{\beta} \pm 1.96 \hat{\sigma}_{\hat{\beta}}$, accept H_0 ; otherwise reject H_0 . On the other hand, if the true variance of β , σ_{β}^2 , were known then the test could be rephrased to read: Accept H_0 if the observed value of $\hat{\beta}$ lies in the interval $\beta \pm 1.96 \sigma_{\beta}$; otherwise reject H_0 .

In econometric work one does not and cannot know the true value of β or any of the other parameters. At the same time the sample data are employed to generate

^{12/} This comparison of alternative tests is conducted by an analysis of their power functions. See: P.G. Hoel, Introduction to Mathematical Statistics, Wiley, New York, Third Edition, 1962, pp. 54-56.

maximum likelihood estimates for all parameters. Consequently, it is not possible to set up and test a null hypothesis of the type described above where specific values are tested against other alternatives. The only null hypothesis it is possible to subject to test is that the true value of the parameter is zero, i.e.

$$H_0: \beta = 0$$

where the alternative hypothesis is $H_1: \beta \neq 0$. Given the estimates $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}$ one can determine for a particular level of confidence, say Type I error of 5%, whether the interval $\hat{\beta} \pm 1.96 \hat{\sigma}_{\hat{\beta}}$ includes $\beta=0$. In other words it is possible to conduct a test to determine whether the true parameter value, β , can be accepted as being equal to zero. Manifestly, if H_0 is accepted then it is necessary to accept the corollary--namely, the variable associated with β can be deleted from the hypothesis.

Even though the null hypothesis, $H_0: \beta=0$, is not rejected by a particular sample estimate, $\hat{\beta}$, it should not be forgotten that the acceptance of H_0 does not rule out the possibility that the data may be consistent with a number of other hypotheses. Indeed, a test of this null hypothesis, within the context of a specific hypothesis and sample data, can only provide information about the probability of the sample estimates coming from a population with a true parameter value equal to zero. For example, if the estimate of the parameter's sample error is large relative to the estimate of the parameter itself the interval $\hat{\beta} \pm 1.96 \hat{\sigma}_{\hat{\beta}}$ may well include the point, $\beta=0$. Under such circumstances a test of the hypothesis, $H_0: \beta=0$, will lead to the acceptance of the null hypothesis. For if $\hat{\beta} = 1.4$ and $\hat{\sigma}_{\hat{\beta}} = 0.8$ the interval would be given by $(-0.17 \text{ to } 2.97)$ which is clearly consistent with the null hypothesis. At the same time, it is also consistent with a large number of other hypotheses including the hypothesis that

the true value of β is greater than one. Unfortunately, in econometric work it is not possible to determine the true value of β . It is therefore not possible to subject the sample estimates to other more revealing tests.

With regard to statistical tests one further point is worth noting. In the beginning of this chapter the assumptions underlying the error term $\underline{u}$ were discussed. These assumptions state in part that $\underline{u}$ is a random variable with a normal distribution which has a mean equal to zero and a constant variance. Within the context of a particular hypothesis, e.g. $y = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + u$, the assumptions about the mean and variance of $\underline{u}$ can be subjected to test. By rearranging the hypothesis $\underline{u}$ can be shown as a function of the remaining variables:

$$u = y - \alpha_0 - \alpha_1 z_1 - \alpha_2 z_2 .$$

The null hypothesis is that $\underline{u}$ is normally distributed with mean equal to zero i.e.

$$f(u, \theta) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{u - \theta}{\sigma} \right)^2}$$

$$H_0: \theta = 0$$

$$H_1: \theta \neq 0$$

To subject this hypothesis to test one would derive the sample estimate $\hat{\underline{u}}$, construct the confidence interval, and examine whether $\underline{\theta} = E[u] = 0$ is contained within this interval. In order to test whether the variance of $\underline{u}$ is constant it would be necessary to generate at least two sample estimates of $\hat{\underline{\sigma}}_u$. The test would then consist of an attempt to determine whether these estimates were consistent with the hypothesis that they came from two normal populations with the same variance. While this test employs the $\underline{F}$ distribution instead of the normal or $\underline{t}$ distribution as above, the procedure is quite straight forward and the null hypothesis is either accepted or rejected as before. In a similar manner it is also possible, although

somewhat more complicated, to test the remainder of the statistical assumptions about $\underline{u}$. Hence, although statistical tests on the individual parameter values of a particular hypothesis are frequently inconclusive it is possible to subject the assumptions about the error term to a number of statistical tests.

Chapter 8

SOME PROBLEMS OF APPLICATION

The previous chapter is concerned with the basic assumptions and measurement techniques that delimit the foundations of econometric analysis. To keep the analysis within reasonable bounds of simplicity, examples of models and statistical hypotheses were employed which were linear in parameters as well as variables. Frequently, however, the econometrician is required to use more complex forms in the statement of his hypotheses. While no attempt will be made to investigate all such deviations from strictly linear models, the purpose of this chapter is to explore some of the problems that confront the economist when he tries to apply these statistical methods to the estimation of a variety of economic relations.

1. Lagged Endogenous Variables

In econometrics it often happens that one wants to represent the behavior of a system where the values of certain variables in the current period depend directly on their values in the previous period. For instance, from studies of the behavior of consumers it may appear that total consumption in one period is a function of consumption in the previous period. Such behavior could also be expressed by a function which noted that the proportion of income spent on consumption in a specific period depends directly on the proportion of income consumed in the previous period. Many theories describing the cyclic behavior of certain aggregate economic variables also have this property.^{1/}

^{1/}See, for example, the well known models of economic growth and trade cycles: R.F. Harrod, "Essay in Dynamic Theory," Economic Journal, Vol. 49, 1939, pp. 14-33; E. Domar, "Capital Expansion, Rate of Growth and Employment," Econometrica, Vol. 14, 1946, pp. 137-147; and N. Kaldor, "A Model of the Trade Cycle," Economic Journal, Vol. 50, 1940, pp. 78-93. Both the Harrod-Domar and the Kaldor models are discussed in some detail in: L.R. Klein, An Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, 1962, Chapter 5.

initial value of $y(t)$ is known and is a constant, i.e. if $y(0) = C$, then the maximum likelihood estimate of β is given by

$$\beta = \frac{\sum_{i=1}^n y_i y_{i-1}}{\sum_{i=1}^n y_{i-1}^2} \quad (8.3)$$

Usually maximum likelihood estimates are unbiased. In this case, however, the estimate of β is biased where the extent of the bias is a function of the sample size and the initial value C . For small samples the bias can be quite large (approximately 25%) and even though an increase in the sample size reduces the bias it cannot be eliminated completely. At the same time, it should not be forgotten that to be able to judge the reliability of a particular estimate it must be unbiased. Hence, a procedure which generates a biased estimate of β cannot be employed if we are to be able to subject the resulting relation to test.

To obtain an unbiased estimate of β it is necessary to limit the sample size to one item. If one begins at the time period, $t=0$, and the value of $y(0)=C$ then the value of the relation is given by, $y(1)=\beta C + u(t)$. Under these conditions an unbiased estimate of β is given by the degenerate least squares estimate.

$$\hat{\beta} = \frac{y(1)}{C} \quad (8.4)$$

Consequently, the general form of the estimating equation is given by

$$\hat{\beta} = \frac{y(t)}{y(t-1)} \quad (8.5)$$

For example, if relation (8.2) refers to the consumption behavior of one individual then it is possible to derive an unbiased (though inefficient) estimate of the parameter β . As long as the values of $y(t-1)$ and $y(t)$ can be observed one can derive such an estimate. But, if (8.2) refers to the aggregate behavior of a group of consumers, then before (8.4) or (8.5) can be employed to estimate β one

has to be able to show that the amount consumed by all of these consumers during period $(t-1)$ is the same. In other words, if during period $(t-1)$ all consumers under investigation can be shown to have consumed exactly $y(t-1)$ dollars worth of commodities each, then β can be estimated by waiting until the next period, observing for one consumer the value of $y(t)$, and computing $\hat{\beta}$ according to (8.5)^{3/}.

Manifestly, if it were possible to control the initial conditions so that one was always assured of the initial equality of the $y(t-1)$'s, then one could average the individual estimates, $\hat{\beta}_i$, and derive an unbiased and efficient estimate of β . Unfortunately, economists working with aggregate data, such as time series, are not able to inspect or control the initial conditions. Further, when working with time series data too short a time interval can produce strong dependencies in the values of several variables between one period and the next. As a result, the presence of lagged endogenous variables (auto regressiveness) can severely restrict the ability to derive unbiased estimates of a model's parameters.^{4/}

2. Simultaneous Interdependence

In section 2.B of the previous chapter the problem was discussed of simultaneously estimating the parameters of several equations. It was noted that a maximum likelihood function can be formed and employed to produce the

^{3/} For further discussion see: S. Valavanis, op. cit., pp. 57-61.

^{4/} For detailed studies of the effects of autoregression in major economic models see: G.H. Orcutt, "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States 1919-1932," Journal of the Royal Statistical Society, sec. B, Vol. 10, 1948, pp. 1-53; and A.J. Gartaganis, "Autoregression in the United States Economy, 1870-1929," Econometrics, Vol. 22, April 1954, pp. 228-243.

which can be represented in its simplest form by the relation

$$y(t) = \alpha + \beta_1 z_1(t) + \beta_2 z_2(t) + \dots + \beta_m z_m(t) + u(t) \quad (8.1)$$

where some of the $z_i(t)$ are lagged values of $y(t)$, e.g. $z_1(t) = \gamma_1(y_{t-1})$, $z_2(t) = \gamma_2(y_{t-2})$, etc. Since $y(t)$ represents the endogenous variable in this relation, then z_1 , z_2 and any of the other z 's which are lagged values of $y(t)$ must also be endogenous variables. For if none of the z 's are lagged values of $y(t)$ then $y(t)$ is the only variable dependent on the value of $u(t)$. But once some of the z 's are lagged values of $y(t)$ then these values are necessarily correlated with some past values of $u(t)$. Accordingly, the lagged variables cannot be independent of the error term, and one of the statistical assumptions about the error term is violated, i.e. $\text{Cov}[u(t), z_{(t-i)}] \neq 0$ for all t and all i .^{2/} The assumptions concerning the error term are employed because once satisfied they permit the techniques of maximum likelihood and least squares (if appropriate) to be used to estimate the unknown parameter values. If one of these assumptions no longer holds then it is reasonable to expect certain difficulties in applying these estimating techniques.

To illustrate the difficulties consider the simple relation

$$y(t) = \beta y(t-1) + u(t) \quad (8.2)$$

In this case $y(t)$ as well as $y(t-1)$ are correlated with the values of the error term $u(t)$ and $u(t-1)$ respectively. Such a situation does not satisfy the independence requirement on the error term. In order to estimate the value of β it is necessary to take a sample of values of both $y(t)$ and $y(t-1)$. If the

^{2/}See Chapter 7, p. 134.

maximum likelihood estimates of the parameters in the general system given by

$$\begin{aligned}
 y_1 + \alpha_{12}y_2 + \dots + \alpha_{1k}y_k + \beta_{11}z_1 + \beta_{12}z_2 + \dots + \beta_{1m}z_m &= u_1 \\
 \alpha_{21}y_1 + y_2 + \dots + \alpha_{2k}y_k + \beta_{21}z_1 + \beta_{22}z_2 + \dots + \beta_{2m}z_m &= u_2 \\
 \vdots & \\
 \alpha_{k1}y_1 + \alpha_{k2}y_2 + \dots + y_k + \beta_{k1}z_1 + \beta_{k2}z_2 + \dots + \beta_{km}z_m &= u_k
 \end{aligned} \tag{8.6}$$

Further, it was pointed out that the value of each parameter depended upon the values of some parameters in the remaining equations. Thus, if each equation were treated as an independent unit and the parameters were estimated by least squares, then the resulting estimates would be at variance with those generated by maximizing likelihood function. Because the maximum likelihood estimates are unbiased, the least squares estimates would clearly be biased and hence of limited value.

Consider the more usual case where the basic model consists of equations including error terms, as well as one or two economic identities. Earlier it was pointed out that if the parameters of an identity are to be estimated it must be transposed into a normal statistical hypothesis.^{5/} For example, if the basic model were given by:

$$\begin{aligned}
 y(t) &= \alpha + \beta z_1(t) + u(t) \\
 y(t) + z_2(t) &= z_1(t)
 \end{aligned} \tag{8.7}$$

the second relation would have to be translated into a statistical hypothesis before estimation procedures could begin. One way of performing this

^{5/} See pp. 143-144.

transposition is to substitute $[(y(t) + z_2(t))]$ for $\underline{z}_1(t)$ in the first equation, and $[\alpha + \beta \underline{z}_1(t) + u(t)]$ for $\underline{y}(t)$ in the second to get

$$\begin{aligned} y(t) &= \alpha + \beta[y(t) + z_2(t)] + u(t) \\ z_1(t) &= z_2(t) + [\alpha + \beta \underline{z}_1(t) + u(t)] \end{aligned}$$

Simplifying these relations we have

$$\begin{aligned} y(t) &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} z_2(t) + \frac{u(t)}{1-\beta} \\ z_1(t) &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} z_2(t) + \frac{u(t)}{1-\beta} \end{aligned} \tag{8.8}$$

Notice, that both relations are now statistical hypotheses with $\underline{z}_2(t)$ being the only exogenous variable. The error term is given by $\frac{u(t)}{1-\beta}$ which has the same properties as $\underline{u}(t)$ except for a shift in its variance. Further, each relation is now independent of the other, i.e. each has a single dependent variable and the same independent variable. Thus, the parameters of each equation can be estimated independently of each other. For a single, linear equation the technique of least squares produces estimates that are identical to maximum likelihood estimates. Consequently, if the parameters of (8.8) are denoted by $\underline{\alpha}' = \alpha/1-\beta$, $\underline{\beta}_1' = \beta/1-\beta$, and $\underline{\beta}_2' = 1/1-\beta$ the parameters of the following two relations can be estimated by least squares:

$$\begin{aligned} y(t) &= \alpha' + \beta_1' z_2(t) + \frac{u(t)}{1-\beta} \\ z_1(t) &= \alpha' + \beta_2' z_2(t) + \frac{u(t)}{1-\beta} \end{aligned} \tag{8.9}$$

Once the estimates of $\underline{\alpha}'$'s, $\underline{\beta}_1'$ and $\underline{\beta}_2'$ are obtained one can immediately compute the estimates $\hat{\alpha}$ and $\hat{\beta}$, since $\hat{\alpha}' = \frac{\hat{\alpha}}{1-\hat{\beta}}$, $\hat{\beta}_1' = \frac{\hat{\beta}}{1-\hat{\beta}}$, and $\hat{\beta}_2' = \frac{1}{1-\hat{\beta}}$.

In order to transpose (8.7) into (8.9) and to estimate the new parameters by least squares three conditions must be met. The first is that the relation between the dependent and independent variables has to be one-to-one. A

many-to-one relation, e.g. $y(t) = \alpha + \beta_1 z_1(t) + \beta_2 z_1^2(t) + u(t)$, allows $z_1(t)$ to have two values for every value of $y(t)$. Unless the y 's and the z_1 's are equally numerous^{6/} the Jacobian is undefined and neither least squares nor maximum likelihood methods can directly be applied. The second condition is that the system of equations must form a set of independent statistical hypotheses. If the first equation in (8.9) contained the variable, $z_1(t)$, or if the second equation contained the variable, $y(t)$, or if both conditions were true (8.9) would no longer consist of two independent relations. Accordingly, least squares would be inappropriate and the parameters would have to be estimated by the maximum likelihood approach.

If the first two conditions are satisfied but the parameters $\alpha', \beta_1', \beta_2', \dots, \beta_n'$ are such that they do not uniquely define the original parameters $\alpha, \beta_1, \beta_2, \dots, \beta_n$, then these equations cannot be estimated either by least squares or maximum likelihood methods. Hence, the third condition requires that the estimates of the new parameters to uniquely determine the estimates of the parameters in the original equations. For example, suppose the original equations are such that some of the transposed parameters are given by:

$$\begin{aligned}\beta_1' &= \beta_1 + \beta_2' \\ \beta_2' &= \beta_2 + \beta_3' \\ 0 &= \beta_2 + \beta_2' \\ 0 &= \beta_2 + \beta_1'\end{aligned}\tag{8.10}$$

From (7.10) it is clear that there are two possible estimates of β_2 i.e.

$\hat{\beta}_2 = -\hat{\beta}_1'$ and $\hat{\beta}_2 = -\hat{\beta}_2'$. In such a situation it is not possible to estimate the values of the parameters. In order to do so the original relations must be altered

^{6/}See Chapter 7, Sec. 2.B.

in such a way that the ambiguity disappears. The problem of ambiguity or non-uniqueness of the parameters is part of a general class of difficulties which unless resolved completely obstructs the process of estimation. These difficulties are encompassed by what is called the identification problem. And it is toward an examination of this class of problems that the next section is directed.

3. The Identification Problem^{7/}

As long as the econometric model consists of a single equation, with one dependent variable represented as a linear function of the parameters of the exogenous variables, the estimation of these parameters is quite straight forward. The minute the model is enlarged to include more than one equation the parameters can be estimated only if these equations are fully identified. A model of two equations, each of which contains one endogenous and two exogenous variables, is fully identified if neither of the two relations, or any two relations which can be derived from them, look alike from a statistical point. That is to say, if each of the two relations is logically independent of the other^{8/} then the system itself is identified. In a similar manner, if the number of equations is increased to n , the model is identified if and only if,

^{7/} An excellent discussion of the identification problem is provided by: T.C. Koopmans, "Identification Problems in Economic Model Construction," in W.C. Hood and T.C. Koopmans (eds), Studies in Econometric Method, Cowles Commission Monograph, No. 14, Wiley, New York, 1953, pp. 27-48.

^{8/} Two axioms, postulates or hypotheses, are independent of each other if neither can be derived from the other. (See Chapter 2, fn.4).

all the model's functional relations are independent of each other.

Consider, for example, a model which states that the quantity demanded of a certain class of items is a function of its price, that the quantity supplied is a function of the market price, and that market equilibrium occurs when the quantity demanded equals the quantity supplied. If y represents quantity and z represents price, this model can be expressed by:

$$\begin{aligned}y_d(t) &= \alpha_1 + \beta_1 z(t) + u_1(t) \\y_s(t) &= \alpha_2 + \beta_2 z(t) + u_2(t) \\y_d(t) &= y_s(t) + u_3(t)\end{aligned}\tag{8.11}$$

where the subscripts d and s refer to the demand and supply relations respectively. For (8.11) to be identified the three relations must be independent of each other. But if the third equation is substituted into the first, (8.11) is transposed into the following system of two relations:

$$\begin{aligned}y_{(s)}(t) &= \alpha_1 + \beta_1 z(t) + u_1(t) - u_3(t) \\y_s(t) &= \alpha_2 + \beta_2 z(t) + u_2(t)\end{aligned}\tag{8.12}$$

If $u_1(t)$, $u_2(t)$, $u_3(t)$ represented observables these two equations would not be statistically identical. However, the u 's represent non-observable, random disturbances. Thus, the term, $u_1(t) - u_3(t)$, is also a random variable and is not statistically distinguishable from the term $u_2(t)$. As a result, if one collected a sample of data and tried to estimate the parameter values, α_1 and β_1 , there would be no statistical way of determining which estimate belonged to which parameter. Since these two relations are statistically indistinguishable the system in (8.11) is not identified.

In order to delimit the necessary and sufficient conditions for identification in econometric models consider the general case:

$$\begin{aligned}
 y_1 & + \beta_{11}z_1 + \beta_{12}z_2 + \beta_{13}z_3 + \beta_{14}z_4 = u_1 \\
 y_2 + \alpha_{23}y_3 + \beta_{21}z_1 & + \beta_{24}z_4 = u_2 \\
 \alpha_{31}y_1 + y_3 + \beta_{31}z_1 + \beta_{32}z_2 + \beta_{33}z_3 + \beta_{34}z_4 & = u_3
 \end{aligned} \tag{8.13}$$

where the y 's represent endogenous variables, the z 's exogenous variables, and the parameters α and β fixed constants. The system (8.13) is identified if the three equations are independent of each other. If the three equations are independent then there is a unique triplet of values (y_1, y_2, y_3) which is the solution to this system.

The necessary and sufficient condition for a unique solution is that the matrix of coefficients of the endogenous variables y be non-singular. This matrix is given by the row and column array of α_{ij} 's. The matrix of α 's, the A matrix, is non-singular if it is a square array of terms which has a non-zero determinantal value.^{9/} If any row in the A matrix can be shown to be a linear combination of the terms of any other row, then the value of the determinant of A is zero and matrix A is singular. Consequently, if the relations in (8.13) are such that the coefficients of the dependent variables form a non-singular matrix, then a unique solution is assured and the system is identified.

One way of investigating the properties of the A matrix, and hence whether a specific set of hypotheses is identified, is to rewrite the equations in terms of the dependent variables (as in (8.8) and (8.9) above).^{10/} For example, the set of relations in (8.13) can be rewritten by a process of appropriate substitutions

^{9/} See Appendix

^{10/} I am indebted for this approach to: S. Valavanis, op. cit., Chapter 6.

into the following form:

$$\begin{aligned} y_1 &= \gamma_{11}z_1 + \gamma_{12}z_2 + \gamma_{13}z_3 + \gamma_{14}z_4 + v_1 \\ y_2 &= \gamma_{21}z_1 + \gamma_{22}z_2 + \gamma_{23}z_3 + \gamma_{24}z_4 + v_2 \\ y_3 &= \gamma_{31}z_1 + \gamma_{32}z_2 + \gamma_{33}z_3 + \gamma_{34}z_4 + v_3 \end{aligned} \quad (8.14)$$

where the γ 's represent the coefficients which result from this transformation and the v 's the random disturbances. Both the γ 's and the v 's are linear combinations of the original coefficients and disturbances in (8.13).

To compute the values of the γ 's one begins with the relations in (8.13) where from the first equation it can be seen that

$$y_1 = -\beta_{11}z_1 - \beta_{12}z_2 - \beta_{13}z_3 - \beta_{14}z_4 + u_1$$

From (8.14)

$$y_1 = \gamma_{11}z_1 + \gamma_{12}z_2 + \gamma_{13}z_3 + \gamma_{14}z_4 + v_1$$

Hence it follows that

$$\begin{aligned} -\beta_{11} &= \gamma_{11} \\ -\beta_{12} &= \gamma_{12} \\ -\beta_{13} &= \gamma_{13} \\ -\beta_{14} &= \gamma_{14} \end{aligned} \quad (8.15)$$

By substituting the value for y_1 from (8.14) into the third equation of (8.13) we have

$$\begin{aligned} y_3 &= -\alpha_{31}(\gamma_{11}z_1 + \gamma_{12}z_2 + \gamma_{13}z_3 + \gamma_{14}z_4) - \beta_{31}z_1 - \beta_{32}z_2 - \beta_{33}z_3 \\ &\quad - \beta_{34}z_4 + u_3 \end{aligned}$$

or

$$\begin{aligned} y_3 &= -(\alpha_{31}\gamma_{11} + \beta_{31})z_1 - (\alpha_{31}\gamma_{12} + \beta_{32})z_2 - (\alpha_{31}\gamma_{13} + \beta_{33})z_3 \\ &\quad - (\alpha_{31}\gamma_{14} + \beta_{34})z_4 + u_3 \end{aligned}$$

From (8.14)

$$y_3 = \gamma_{31}z_1 + \gamma_{32}z_2 + \gamma_{33}z_3 + \gamma_{34}z_4 + v_3$$

Consequently the values of β are given by:

$$\begin{aligned} -\beta_{31} &= \alpha_{31}\gamma_{11} + \gamma_{31} \\ -\beta_{32} &= \alpha_{31}\gamma_{12} + \gamma_{32} \\ -\beta_{33} &= \alpha_{31}\gamma_{13} + \gamma_{33} \\ -\beta_{34} &= \alpha_{31}\gamma_{14} + \gamma_{34} \end{aligned} \tag{8.16}$$

Similarly by substituting the value for y_3 from (8.14) into the second equation of (8.13) we get

$$y_2 = -\alpha_{23}(\gamma_{31}z_1 + \gamma_{32}z_2 + \gamma_{33}z_3 + \gamma_{34}z_4) - \beta_{21}z_1 - \beta_{24}z_4 + u_2$$

or

$$y_2 = -(\alpha_{23}\gamma_{31} + \beta_{21})z_1 - \alpha_{23}\gamma_{32}z_2 - \alpha_{23}\gamma_{33}z_3 - (\alpha_{23}\gamma_{34} + \beta_{24})z_4 + u_2$$

From (8.14)

$$y_2 = \gamma_{21}z_1 + \gamma_{22}z_2 + \gamma_{23}z_3 + \gamma_{24}z_4 + v_2$$

Accordingly, the remaining values of β are given by

$$\begin{aligned} -\beta_{21} &= \alpha_{23}\gamma_{31} + \gamma_{21} \\ 0 &= \alpha_{23}\gamma_{32} + \gamma_{22} \\ 0 &= \alpha_{23}\gamma_{33} + \gamma_{23} \\ -\beta_{24} &= \alpha_{23}\gamma_{34} + \gamma_{24} \end{aligned} \tag{8.17}$$

Having determined the values of the β 's in terms of the parameters α and γ it is now possible to assess whether each relation in the original system (8.13) is identified. From (8.15) it is clear that there is a one-to-one relation between these β 's and γ 's. So that a knowledge of the values of--i.e. estimates of-- $\gamma_{11}, \gamma_{12}, \gamma_{13}$ and γ_{14} , uniquely determines the values of $\beta_{11}, \beta_{12}, \beta_{13}$ and β_{14} . Accordingly, these four parameters ($\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}$) are exactly identified.

In (8.16), however, there are four equations in five unknowns. That is to say, estimates of the values of γ_{11} , γ_{12} , γ_{13} , γ_{14} , γ_{31} , γ_{32} , γ_{33} and γ_{34} will not permit the values of the unknowns β_{31} , β_{32} , β_{33} , β_{34} and α_{31} to be determined. The system of equations is underdetermined, and as a result the parameters β_{31} , β_{32} , β_{33} , β_{34} and α_{31} are underidentified.

The equations in (8.17) present yet another problem. Here the two middle relations state that the value of α_{23} is determined in two different ways, i.e.

$$\alpha_{23} = -\gamma_{22}/\gamma_{32} \quad \text{and} \quad \alpha_{23} = -\gamma_{23}/\gamma_{33}$$

Unless it is specified that the value of $\gamma_{22} = \gamma_{23}$ and $\gamma_{32} = \gamma_{33}$ it is clear that the equations in (8.17) over-determine the value of α_{23} . As a result the parameter α_{23} is over-identified.

In the previous section it was noted that the equations of a system in the form of (8.14) can be independently estimated by least squares if three conditions are satisfied. The first condition requires a one-to-one correspondence between the endogenous and exogenous variables. The second states that the equations must be statistically independent of each other. And the third requires that the parameters of (8.14) uniquely define the parameters of the original system (8.13). If these conditions are met then it is possible to derive least squares estimates of the γ 's. Once the γ 's are estimated it follows that one can immediately specify the maximum likelihood estimates of the α 's and the β 's. But, from the analysis of the relations between the parameters of (8.13) and (8.14) it is clear that the system in (8.13) does not meet all of these conditions. In particular, the relations in (8.16) and (8.17) demonstrate that some of the parameters e.g., β_{31} , β_{32} , β_{33} , β_{34} , and α_{31} are underidentified while α_{23} is overidentified. Therefore, it is not possible to take the system

of equations in (8.14) and from least squares estimates of these parameters derive estimates for the parameters of the original system in (8.13).

To estimate the parameters of (8.13) the equations must be identified. Although there are several techniques by which an unidentified system can be transformed so that it is fully identified, there are two general rules by which the identification of individual equations can be determined. The first is a necessary condition for identification while the second is both necessary and sufficient.

The first rule is concerned with the statistical independence of each equation. If all the equations of a particular model are of the endogenous and exogenous variables of the system then these equations are not going to be statistically independent. If each equation has a certain number of the total set of variables absent from it then it is possible that these equations are independent of each other. Consequently, to determine the identifiability of an equation it is the variables that are absent from it which become the critical factor. Accordingly, it is not surprising that the first, necessary condition is stated in terms of the variables which are absent from a particular equation: If an equation is to be exactly identified it is necessary that the number of variables absent from it be equal to the number of dependent variables minus one.

Employing the customary econometric notation a variable present in a specific equation is labelled with an asterisk, while a variable of the system that is absent from this equation is denoted by two asterisks. Hence, the first equation of (8.13) can be written:

$$y_1^* + y_2^{**} + y_3^{**} + \beta_{11}z_1^* + \beta_{12}z_2^* + \beta_{13}z_3^* + \beta_{14}z_4^* = u_1 \quad (8.18)$$

Since there are three endogenous variables in (8.13) the necessary condition for the identifiability of this equation is that (8.18) must contain two, double asterisk variables. In fact (8.18) meets this condition. Using the same notation the second and third equations of (8.13) are given by:

$$y_1^{**} + y_2^* + \alpha_{23}y_3^* + \beta_{21}z_1^* + z_2^{**} + z_2^{**} + \beta_{24}z_4^* = u_2 \quad (8.19)$$

$$\alpha_{31}y_1^* + y_2^{**} + y_3^* + \beta_{31}z_1^* + \beta_{32}z_2^* + \beta_{33}z_3^* + \beta_{34}z_4^* = u_3 \quad (8.20)$$

In (8.19) there are three, double asterisk variables. But, to satisfy the necessary condition there should only be two variables absent from it.

Similarly, (8.20) also fails to satisfy the necessary condition as it contains only one, double asterisk variable.

One way of altering (8.19) and (8.20) to meet this requirement is to add or subtract the appropriate number of variables by setting the relevant parameters to non-zero values. In (8.19) variables z_2 and z_3 are absent and hence their parameter values were originally specified to be equal to zero. If this decision is changed and either z_2 or z_3 is given a non-zero parameter value, e.g. β_{22} or β_{23} are included in (8.19), then (8.19) will satisfy the necessary condition for identification. In the same fashion, (8.20) will satisfy this requirement if one of the parameters, α_{31} , β_{31} , β_{32} , β_{33} , β_{34} , are declared to be equal to zero. Such adjustments obviously change the original set of hypotheses in (8.13). But without making these alternations the system is unidentified and as a result it is not possible to estimate and test the original hypotheses recorded in (8.13).

The second rule approaches the problem of identification from the relations between the parameters of (8.13) and (8.14). As has already been noted, if the parameters of (8.14) uniquely determine the parameters of (8.13), then the original system of equations is identified. In all other cases, i.e. if the parameters of (8.14) over- or under-determine the parameters of (8.13), the system is over- or under-identified. Manifestly, the unique determination of the parameters of (8.13) is a function of the presence or absence of the appropriate parameters in the equations of (8.13) itself. Consequently, once again the criterion which governs identification is stated in terms of those parameters which are absent from a particular equation: If an equation is to be identified it is both necessary and sufficient that the matrix of parameters $\underline{C}^{**}$ (which is formed by deleting the columns from the two matrices $\underline{A}$ and $\underline{B}$ which correspond to the variables present in the relevant equation) has the rank equal to the number of endogenous variables minus one.

In (8.13) the matrix $\underline{C}$ represents the total matrix of parameter values, e.g.

$$\underline{C} = \begin{bmatrix} 1 & 0 & 0 & \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & 1 & \alpha_{23} & \beta_{21} & 0 & 0 & \beta_{24} \\ \alpha_{31} & 0 & 1 & \beta_{31} & \beta_{32} & \beta_{33} & \beta_{34} \end{bmatrix}$$

Consider the first equation. To form the matrix $\underline{C}^{**}$ one deletes from $\underline{C}$ all columns which correspond to variables in this equation, i.e. one deletes all columns in which there are non-zero entries in the first row of $\underline{C}$. For the first equation of (8.13), $\underline{C}_1^{**}$ is given by:

$$\underline{C}_1^{**} = \begin{bmatrix} 0 & 0 \\ 1 & \alpha_{23} \\ 0 & 1 \end{bmatrix}$$

Similarly for the second equation:

$$\underline{C}_2^{**} = \begin{bmatrix} 1 & \beta_{12} & \beta_{13} \\ 0 & 0 & 0 \\ \alpha_{31} & \beta_{32} & \beta_{33} \end{bmatrix}$$

And for the third equation of (8.13)

$$\underline{C}_3^{**} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The rule states that (8.13) is identified if each of the matrices, $\underline{C}_1^{**}$, $\underline{C}_2^{**}$, and $\underline{C}_3^{**}$, has a rank equal to the number of endogenous variables minus one. In this case, the rank must be equal to 2. A matrix has rank r if at least one of the matrix's sub-matrices is a square array, r x r, which has a non-zero determinant, and if all remaining square sub-matrices of higher order have determinantal values equal to zero.

For example, $\underline{C}_1^{**}$ is a 2 x 3 matrix. The largest square array is 2 x 2.

In $\underline{C}_1^{**}$ there are three such sub-matrices

$$\begin{bmatrix} 0 & 0 \\ 1 & \alpha_{23} \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & \alpha_{23} \\ 0 & 1 \end{bmatrix}$$

The first two have a determinantal value equal to zero. But the last has a non-zero determinant. Hence, the rank of $\underline{C}_1^{**}$ is 2, and the first equation

is identified. $\underline{C}_2^{**}$ is a square 3 x 3 array whose determinantal value is zero.

However, $\underline{C}_2^{**}$ is composed of nine sub-matrices of order 2 x 2, including $\begin{bmatrix} 1 & \beta_{12} \\ \alpha_{13} & \beta_{32} \end{bmatrix}$

which has a non-zero determinant. Hence, $\underline{C}_2^{**}$ also has a rank equal

to 2. $\underline{C}_3^{**}$, on the other hand, is a 1 x 3 matrix. Its only type of square

sub-matrix is a 1×1 matrix which precludes $\underline{C}_3^{**}$ from having a rank equal to 2. As a result, by this criterion the first two equations are identified, while the third is not. Consequently, the system given in (8.13) is unidentified.

In order to satisfy the first criterion, a necessary condition for identification, one solution was to add one variable to the second equation and delete one from the third. One such possible arrangement is given by

$$\begin{aligned} y_1 &+ \beta_{11}z_1 + \beta_{12}z_2 + \beta_{13}z_3 + \beta_{14}z_4 = u_1 \\ y_2 + \alpha_{23}y_3 + \beta_{21}z_1 + \beta_{22}z_2 &+ \beta_{24}z_4 = u_2 \\ \alpha_{31}y_1 + y_3 + \beta_{31}z_1 &+ \beta_{33}z_3 + \beta_{34}z_4 = u_3 \end{aligned} \quad (8.21)$$

From (8.21) the $\underline{C}^{**}$ matrices are now given by

$$\underline{C}_1^{**} = \begin{bmatrix} 0 & 0 \\ 1 & \alpha_{23} \\ 0 & 1 \end{bmatrix}$$

$$\underline{C}_2^{**} = \begin{bmatrix} 1 & \beta_{13} \\ 0 & 0 \\ \alpha_{31} & \beta_{33} \end{bmatrix}$$

$$\underline{C}_3^{**} = \begin{bmatrix} 0 & \beta_{12} \\ 1 & \beta_{22} \\ 0 & 0 \end{bmatrix}$$

Since $\underline{C}_1^{**}$ is unaffected by the alternations in (8.21) it still has a rank equal to 2. The addition and subtraction of parameters has affected $\underline{C}_2^{**}$ and $\underline{C}_3^{**}$, and it can readily be shown that these two matrices also have a rank equal to 2. Because the second criterion has been satisfied one can conclude that the adjustments made to meet the first requirement are in this case sufficient to guarantee the identifiability of the new system given in (8.21).

This is not to say that any alteration which satisfies the first requirement will automatically meet the second. Nor is it being suggested that the only solution is to add and delete the appropriate parameters from the system. Clearly, there are a number of possible ways by which the system in (8.13) can be altered so that its equations are identified. One alternative, noted in the discussion of the relations in (8.17), is to set some of the parameters in the reduced form (8.14) equal to each other. But whatever adjustment is chosen in the initial specification of the system's hypotheses, if the system fails to satisfy either of these two criteria then its parameters can be estimated as the equations are unidentified.

4. Forecasting

Once a model has been estimated from the data of a particular time period it is frequently employed to generate forecasts for the next or later time periods. To examine the way in which the estimated relations can be used to forecast or predict, consider the simple linear relation:

$$y(t) = \alpha + \beta z(t) + u(t) \quad (8.22)$$

This equation is identified since it is linear in both parameters and variables and since there is a one-to-one relation between the enogenous and exogenous variables. To estimate the parameters, α and β , a sample of data on $y(t)$ and $z(t)$ is collected and the maximum likelihood estimates are generated by solving the estimating equations:

$$\begin{aligned} \sum_{i=1}^n (y_i - \alpha - \beta z_i) &= 0 \\ \sum_{i=1}^n (y_i - \alpha - \beta z_i) z_i &= 0 \\ \frac{1}{n} \sum_{i=1}^n (y_i - \alpha - \beta z_i)^2 &= \sigma_n^2 \end{aligned} \quad (8.23)$$

From the solution of these relations the estimated equation for the particular time period under consideration is given by:

$$y(t) = \hat{\alpha} + \hat{\beta} z(t) \quad (8.24)$$

Note that the estimated relation (8.24) no longer contains an error term. This is due to the fact that the expected value of the error term, $u(t)$, is zero. Further, the solution of the relations in (8.23) provides an estimate of the standard error $\hat{\sigma}_u$. The standard error can be used to determine estimates of the standard error of the parameter estimates $\frac{\hat{\sigma}}{\hat{\alpha}}$ and $\frac{\hat{\sigma}}{\hat{\beta}} \cdot \frac{11}{\beta}$. Once these two estimates are formed, it is then possible to construct confidence intervals about the parameter estimates and adjudge their reliability.

In order to forecast the value of y for the next time interval, $t+1$, all that is required is to insert in (8.24) the observed value for $z(t+1)$ and compute the value of $y^F(t+1)$. This procedure assumes, of course, that there have been no significant structural changes and that the estimates $\hat{\alpha}$ and $\hat{\beta}$ form a valid basis for the forecast. If certain changes are observed during the interval, $t+1$, such that (8.22) is no longer appropriate, then it is clear that (8.24) cannot be used for forecasting. In this case a new hypothesis would have to be constructed and its parameters estimated. If this process were carried out during the period $t+1$ then the new estimated relation could be employed to forecast the values of the relevant variables in $t+2$. However, for the purposes of this discussion assume that significant changes have not occurred and that (8.24) is an appropriate forecast relation.

Even though (8.24) is of the simplest possible form it has forecasting characteristics in common with the most complex models. In particular, by employing (8.24) or any other estimated relation to forecast variable values

^{11/} For details of this procedure see Chapter 7, p. 152.

point forecasts are generated. For instance, in period $t+1$ (8.24) provides the point forecast of $\underline{y}^F(t+1)$. By itself the point forecast $\underline{y}^F(t+1)$ can be compared to the observed value of $\underline{y}$ in the period $t+1$. But without specifying some interval about $\underline{y}^F(t+1)$ such a direct comparison does not provide much information. Since $\underline{y}(t)$ and $\underline{z}(t)$ are random variables, $\underline{y}^F(t+1)$ and $\underline{y}(t+1)$ are also random variables. Hence to compare, i.e. test for equality, the values of two random variables one needs to be able to assess whether the value of one, say $\underline{y}(t+1)$, falls within some expected interval about the other, $\underline{y}^F(t+1)$. When dealing with point estimates of parameter values it is possible to construct a confidence interval about these estimates and employ this interval as the basis for testing certain null hypotheses.^{12/} Accordingly, to be able to submit the point forecasts to a similar testing procedure an interval about the forecasted values must be defined.

The theory of measuring the reliability of point forecasts is relevant to the case where the endogenous variable is a linear function of the exogenous variables. If the endogenous variable is normally distributed with a true mean of $\underline{\mu}$ and a true variance of $\underline{\sigma}^2$, tolerance limits are formed by adding and subtracting from the mean, $\underline{\mu}$, a specific multiple $\underline{K}$ of the standard deviation, $\underline{\sigma}$. In other words, if $\underline{\mu}_y$ and $\underline{\sigma}_y$ are the true mean and variance of the distribution, the tolerance limits for a value of the endogenous variable are given by

$$\mu_y \pm K \sigma_y \quad (8.25)$$

where $\underline{K}$ is a parameter and depends on the sample size and the proportion of

^{12/} See Chapter 7, pp. 150-151.

values of y to be included in the interval.^{13/}

Tolerance limits are not the same as confidence limits. A confidence limit for the mean, e.g. $\bar{x} \pm 1.96\sigma$, states that 95% of the time we expect the true population mean to lie between the limits $\bar{x} - 1.96\sigma$ and $\bar{x} + 1.96\sigma$, where $\bar{x}$ represents the sample mean. A 95% tolerance limit on the other hand, states that we expect 95% of the sample values of the endogenous variable to include a proportion p of the values in its distribution. That is to say, the endogenous variable has a distribution (assumed to be normal in this case) from which each sample value, $y(t)$, is derived. For a given level of confidence, say 95%, the tolerance interval delimits the minimum proportion of these sample values that we can expect to be included within the limits. Further by increasing the value of K the probability that the tolerance interval contains at least p of the population can be made to be as close to 1 as desired.

In normal circumstances the true mean and variance of the distribution of the endogenous variable are not known. Accordingly, to construct a tolerance interval it is necessary to employ sample estimates. The sample estimate of the mean is given for (8.24) by the forecast value of y , i.e. $y_{(t+1)}^F$. The estimate of the variance can be computed from the previously generated maximum likelihood estimates so that the tolerance interval is given by:

$$y_{(t+1)}^F \pm K \hat{\sigma}_{y_{(t+1)}^F} \quad (8.26)$$

^{13/} Values of K for normal distributions with constant mean are given in tables by A.H. Bowker, "Tolerance Limits for Normal Distributions," in C. Eisenhart, M. Hastay, and W.A. Wallis, (eds), Techniques of Statistical Analysis, McGraw-Hill, 1947, pp. 102-107.

Consequently, (8.26) defines a range of values of $y_{(t+1)}^F$ within which at least a proportion $\underline{p}$ of the non-sampled observations are expected to fall with a certain probability. As noted above, the value of $\underline{K}$ depends on the proportion $\underline{p}$ of future observations which are to lie in this interval, the probability with which this is to occur, and on the size of the sample. Hence, for a specific sample size, for assigned values of $\underline{p}$, and for the probability of this occurring, say 95% of the time, $\underline{K}$ is determined and an interval for $y_{(t+1)}^F$ is specified.

So far the reliability of a forecast has been considered for a single linear equation like (8.22). In order to examine the case of a model with several linear equations, consider the identified system given in (8.21). As mentioned above, estimates of the parameters of this system can be derived directly by the maximum likelihood method. However, it is also possible to transpose (8.21) into the reduced form:

$$\begin{aligned}y_1 &= \gamma_{11}z_1 + \gamma_{12}z_2 + \gamma_{13}z_3 + \gamma_{14}z_4 + v_1 \\y_2 &= \gamma_{21}z_1 + \gamma_{22}z_2 + \gamma_{23}z_3 + \gamma_{24}z_4 + v_2 \\y_3 &= \gamma_{31}z_1 + \gamma_{32}z_2 + \gamma_{33}z_3 + \gamma_{34}z_4 + v_3\end{aligned}\tag{8.27}$$

and estimate the new parameters, γ_{ij} , by least squares. Since all the equations of (8.27) represent one endogenous variable as a linear function of exogenous variables, and since each of these equations is independent of each other, the relations in (8.27) are in the appropriate form for forecasting. To construct tolerance intervals for the endogenous variables it is necessary to compute the appropriate sample variances and covariances. Once these values are determined, however, the forecasting procedures are the same as for the case of the single equation model outlined above.^{14/}

^{14/}For further detail in forecasting technique see: L.R. Klein, A Textbook of Econometrics, op. cit., pp. 249-276.

Chapter 9

THE EMPIRICAL CONTENT OF ECONOMETRIC THEORY

We, as economists, require testable theories of economic behavior. Without such theories economics as a discipline can never provide scientific explanations or predictions of the many important and interesting, observable economic phenomena. In the examination of classical economic theory it was noted that the testability of a theory depends on the empirical content and testability of the theory's hypotheses. In brief, a theory can be corroborated by test if and only if at least one of its constituent hypotheses can be subjected to a process of refutation by empirical test. To submit a single hypothesis to test it must be possible to observe that the initial conditions are empirically true. In Chapters 5 and 6 all hypotheses within classical economic theory were shown to contain unobservable equilibrium conditions as a part of their initial conditions. As a result, none of these hypotheses can be confuted by empirical test. The object of this chapter is to examine the empirical content, and as a consequence the testability, of econometric hypotheses. The primary goal, of course, is to discover a set of economic hypotheses that can be subjected to a process of refutation by empirical test.

1. The Initial Conditions

A. The Error Term

In order to test an econometric hypothesis it must be possible to ascertain that the relevant initial conditions are empirically true. Since all econometric hypotheses are stochastic, the initial conditions which must be empirically established are those concerned with the error term. In Chapter 7,

one particular set of assumptions surrounding the error term are discussed. Estimating procedures are a function of the assumptions about the error term. Also the most common estimation criterion is that of maximum likelihood. Hence, these assumptions about the error term, while not universally employed by econometricians, are those which are most commonly used. Accordingly, although a different set of assumptions would require a separate analysis of their empirical content, the method by which their content would be determined would be similar to that which is described below.

To facilitate the analysis, consider the error term within the context of a specific demand function:

$$y_{(t)} = \alpha + \beta p_{(t)} + u_{(t)} \quad (9.1)$$

where $y_{(t)}$ is the quantity of a certain commodity that is demanded during period t , $p_{(t)}$ is the prevailing market price, α and β are the parameters to be estimated, and $u_{(t)}$ is the value of the error term for this period. Now the error term is assumed to be a random variable, so that its value at each period of time is a function of its probability density function. Thus, the particular value of u during period t , i.e. $u_{(t)}$, is determined by the probability of this value occurring, which in turn is defined by the density function which describes the total population of the values of u . This population of values of u is assumed to be normally distributed with a mean of zero. Accordingly, if a large sample of values of u were collected one would expect these individual values to describe the outlines of a normal population with zero mean. The assumption of the normal distribution is clearly one of the basic postulates about the error term. Thus, it is pertinent to inquire whether this assumption can be checked.

The answer, of course, is obvious to anyone familiar with the application of statistics to problems of this sort. But, the more interesting point concerns the data that must be collected if this assumption is to be tested. At present the analysis is concerned with (9.1), a simple, linear hypothesis about the demand for a particular commodity with a given time period, t . In order to test the postulate of normality of the error term one can proceed in a number of directions. The value of $\underline{u}(t)$ is the result of a particular sample of data on $\underline{y}(t)$ and $\underline{p}(t)$. To determine the distribution from which $\underline{u}(t)$ is derived one needs a sample of values of $\underline{u}(t)$. For a sufficiently large number of samples of values of $\underline{u}(t)$ it is immediately known that the mean of these samples is normally distributed.^{1/} The problem here, however, is to determine the distribution of the underlying and unobservable population from which these values are derived. That is to say, it is necessary to test the postulate on the basis of actual sample values and not sample means. In fact, there are a number of ways in which this postulate can be checked. One such method is to employ the Kolmogorov-Smirnov statistic. Under the null hypothesis that the population is normally distributed with mean zero it is possible to determine the cumulative distribution of the sample values themselves and test directly the null hypothesis. Another approach is to use the Chi-Square statistic to test the expected frequencies under the same null hypothesis against the observed frequencies from the sample values.^{2/} Whatever the method it is clear that a sample of values of $\underline{u}(t)$ is required.

^{1/}This result is a consequence of the Central Limit Theorem, see: A.M. Mood and F.A. Graybill, op. cit., pp. 149-153.

^{2/}For other methods see any advanced text on statistics, e.g. A.M. Mood and F.A. Graybill, op. cit.; or P.G. Hoel, op. cit.

These values must be collected during period t . Although the minimum sample size varies with the statistic employed, each testing procedure requires a number of estimates of (9.1) to be generated within the relevant time period.

The next assumption about the error term is that the density function does not change with time. That is to say, during period t the value of $u_{(t)}$ comes from a normal population with a mean of zero and a specific variance σ_t^2 . If the density function does not vary with time, then the variance σ_t^2 must remain constant over time. Consequently, one is now concerned with testing for the observed sample values of u_t over several intervals of time. To determine whether the true population variance remains constant over time, it is once again necessary to adopt the null hypothesis that it does and to test against this null hypothesis with the sample data. To perform this test a sample of values of u_t are drawn for a number of time intervals. The sample variances are computed and are used to derive a value for the likelihood ratio λ . Since $-2\log \lambda$ is approximately distributed as the Chi-Square the value of this statistic, for the relevant degrees of freedom and level of confidence, is compared with that derived from the logarithm of the likelihood ratio.^{3/}

Notive once again that the test requires a sample of values of the error term to be gathered for each time interval. Since these data have to be collected to test for the constancy of the true variance over time, it is clear that the same data can be employed to determine whether the true

^{3/}For a complete statement of this test see: A.M. Mood and F.A. Graybill, op. cit., pp. 284-290; and P.G. Hoel, op. cit., pp. 225-228.

population density function is normal with a zero mean.

If each of the sample values of $\underline{u}_{(t)}$ is generated from a random sample of the total population of $\underline{y}$'s and $\underline{p}$'s, then each value of $\underline{u}_{(t)}$ is a random variable. Concurrently, because the values of $\underline{u}_{(t)}$ are arrived at by independent random samples, the sample values of $\underline{u}_{(t)}$ will normally be statistically independent of each other. The only time when this result will not occur is when the value of $\underline{u}_{(t)}$ is directly related to the value of the independent variable, $\underline{p}_{(t)}$. Hence, in the case of the simple linear relation (9.1) it is sufficient to test the fifth and sixth assumptions by only submitting the sixth to an empirical check.

The sixth assumption requires the covariance of $\underline{u}_{(t)}$ and $\underline{p}_{(t)}$ to be zero,^{4/} where the covariance is the first product moment of the observed values about their means, i.e.

$$\text{Cov}(u, p) = E[u - E[u]][p - E[p]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(p - E[p]) f(u, p) dp_{(t)} du_{(t)}$$

Since the joint density of $\underline{u}$ and $\underline{p}$, $f(u, p)$ may well be unknown, the test can be conducted by computing the correlation coefficient between the sample values of $\underline{u}_{(t)}$ and $\underline{p}_{(t)}$. It must not be forgotten that the sixth assumption requires that $\text{Cov}(u_{(t)}, p_{(t-1)}) = 0$ for all $\underline{t}$ and all $\underline{i}$. Consequently, over the relevant number of time periods all correlation coefficients have to be shown to be statistically indistinguishable from zero.

So far the discussion has been concerned with describing the procedures by which one can determine whether the initial conditions implied by the error

^{4/}See Chapter 7, p. 134.

term $u(t)$ in (9.1) are empirically true or false within the context of a particular case. It follows from this analysis that sufficient statistics exist to permit such testing to take place. As a result, the initial conditions surrounding the error term can indeed be put to statistical test. Whether they are in fact satisfied for a specific relation and by a particular set of data is not of interest here. The point to note is that these conditions can be put to statistical test.

The minute one progresses from a single to a many equation model the testing of the assumptions about the error terms becomes somewhat more complex. Instead of dealing with a single error term, one is now concerned with a vector of error terms. Instead of a univariate normal distribution with zero mean, the underlying population is assumed to be a multivariate, or a jointly normal, distribution with a mean of zero. If the multivariate normal is to have a constant variance over successive intervals of time, then the covariances between the respective individual error terms must be constant over time. In an analogous manner the fifth and sixth assumptions require that both the covariances between the error terms and the values of the independent variables be equal to zero.

While the statistical tests are somewhat complicated by the introduction of the multivariate normal, the complications do not preclude the possibility of conducting them. Clearly, the introduction of a many equation model increases both the problem of data collection as well as that of computation. Indeed, if the number of equations is large enough these practical problems may well prevent the testing from taking place. The problem, however, is not whether these tests can in fact be carried out in every imaginable case. This

is not an important point. The question at issue is whether it is possible given as much time, computational facilities, and data as needed to empirically establish within the context of a particular case the presence or absence of these six initial conditions. If these conditions are shown to be statistically true, the one can proceed to test the empirical relevance of the hypothesis or model itself. If the conditions are, within the context of a specific case, statistically false then there is no need to proceed any further. In either event, however, there is no question that there are explicit procedures for ascertaining the statistical truth or falseness of these initial conditions. Consequently, whether these conditions apply to each and every case or not, it is possible to submit them to empirical scrutiny--that is, the initial conditions are members of the class of observable items.

B. The Remaining Conditions

One point which has not yet received proper attention concerns the source or origin of econometric hypotheses. If each time an hypothesis or model is developed it is constructed de novo then the total number of initial conditions are circumscribed and are defined by the hypothesis or model itself. On the other hand, if the hypothesis is suggested by, or is a product of, some prior theoretical framework then it may well occur that parts of this theory are reflected as conditions surrounding the hypothesis. In such a case these conditions then become a part of the initial conditions which must be observed before the hypothesis can be submitted to test.

Consider, for example, the demand relation noted in the previous section

$$y(t) = \alpha + \beta p(t) + u(t) \quad (9.1)$$

Where does this hypothesis come from? And what additional conditions accompany it?

Demand curves are, as has already been shown,^{5/} derived from an analysis of the choice behavior of consumers under equilibrium conditions. If the consumer maximizes his utility function subject to his budget constraint, the demand relation is a direct consequent of this theory of choice. Further, this relation between the price of a commodity and the quantity purchased only refers to those cases where the consumer's taste, income, as well as the prices of related goods, and other environmental factors are both given and unchanging. Not only must all equilibrium conditions be satisfied before the relation can be deduced, but all relevant variables in the environment must remain the same if the relation is to hold at all.

It does not follow that the conditions surrounding the demand relation in classical theory need to be a part of the econometric formulation. Manifestly, the relation (9.1) can be the plain, explicit statement that the quantity purchased of a certain commodity is related to its market price during a specific time interval in the stated manner. Further, that this relation is based upon repeated observations of the behavior of the price and quantity purchased of this commodity. That is to say, over a certain interval of time a sample of values have been observed and recorded and (9.1) represents the best fitting, stochastic relation as in Figure 1. In this case there is no reference to the theory of consumer choice, to the equilibrium conditions, or

^{5/}See Chapters 4 and 5, sections 1 and 1.A respectively.

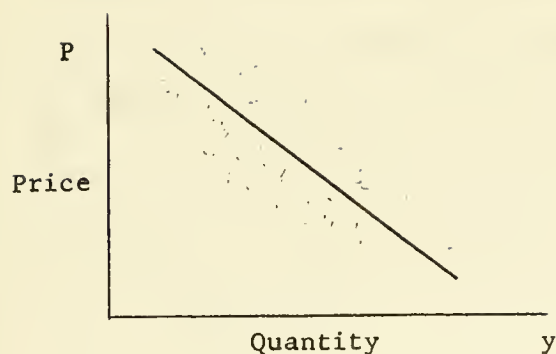


FIGURE 1

to other environmental factors. The relation is based on actual occurrences and the parameters, α and β , are estimated from these sample values. If as in Figure 1 the relation has a negative slope, i.e. β has a negative value, then this is a fact about this sample of values and not a consequent of a theory of consumer behavior.

This is an important point and one to which reference will be made again in the discussion on testing econometric hypotheses. For, if the relation is derived from classical theory, and if the econometric statement of its is merely one way of placing it into testable form, then the initial conditions must include the unobservable equilibrium conditions as well as the ceteris paribus clause. If these are a part of the initial conditions, then the hypotheses cannot be submitted to test. For, although it is possible to statistically test the assumed properties of the error term, it is not possible to show that the equilibrium conditions are satisfied.

An example of this point is offered in a recent discussion of the Cobb-Douglas production function.^{6/} The Cobb-Douglas function is given by

$$P = b L^k C^{1-k} \quad \underline{7/} \quad (9.2)$$

where $\underline{P}$ represents output, $\underline{L}$ labor input, $\underline{C}$ capital input, and $\underline{b}$ and $\underline{k}$ are

^{6/} H.A. Simon and F.L. Levy, "A Note on the Cobb-Douglas Function," Review of Economic Studies, Vol. XXX, June 1963, pp. 93-94.

^{7/} In keeping with these author's presentation the error term is ignored.

parameters of the relation. This equation is a part of the classical theory of production where the partial derivative of output with respect to labor input, $\frac{\partial P}{\partial L}$, is the marginal productivity of labor, and where the parameter $\underline{k}$ represents labor's fraction of total output. Setting the marginal productivity equal to the competitive wage, the result is the relation:

$$\frac{\partial P}{\partial L} = k \frac{P}{L}$$

or
$$k = \frac{\partial L}{\partial P} \frac{L}{P} \quad (9.3)$$

where for a given period of time $\frac{\partial L}{\partial P}$ is a constant.

A number of empirical tests have been conducted to measure the value of $\underline{k}$. One set of tests were concerned with fitting (9.2) directly to sample data on the values of $\underline{P}$, $\underline{L}$, and $\underline{C}$. A second sample of data was then employed to derive a value of $\underline{k}$ from a direct assessment of labor's share of total income. These two values of $\underline{k}$ were observed to be in fair agreement with one another. Consequently, it was inferred that this test corroborated the assumptions underlying (9.2).

Simon and Levy suggest, however, that approximately the same values of $\underline{k}$ will be obtained if the production function is given by the simple linear relation

$$P = a L + d C \quad (9.4)$$

where $\underline{a}$ and $\underline{d}$ are parameters representing the average wage and yield on capital respectively. Labor's fraction of output (income) is introduced as

$$K' = \frac{aL}{P}$$

By dealing with average values of output and labor input, $\bar{P}$ and $\bar{L}$, the value of $\underline{K}$ is given by

$$K = a \frac{\bar{L}}{\bar{P}} \quad (9.5)$$

The issue at hand is not which of these two relations, (9.2) or (9.4), is the "right" one. Rather the point revolves around the question of what it means to have the values of $\underline{K}$ be approximately the same in both of these cases. That is to say, does the fact that the fitted value of $\underline{K}$, from (9.2), agrees with the observed value of $\underline{K}$ corroborate the underlying hypothesis and assumptions of (9.2)? The answer, of course, is no. If the assumptions and initial conditions are enumerated, and observed to be empirically true during the period in which the data are collected, then the evidence on $\underline{K}$ would indeed serve to support these assumptions. But in fact, some of these initial conditions are the ubiquitous, unobservable equilibrium conditions. Accordingly, no claim can be made that the evidence on $\underline{K}$ corroborates these unobservable assumptions.

At the risk of exhausting the reader's patience this argument can be further clarified if the logical notation introduced in Chapter 2 is employed. Let $\underline{Q}$ represent the general theory of production with its concomitant equilibrium conditions, $\underline{R}$ the Cobb-Douglas production function in the form of (9.2), and $\underline{S}$ the consequent of $\underline{R}$ which is the relation denoted by (9.3). If the production function is considered a part of the classical theory of production then the chain of inference is represented by the proposition $\underline{Q} \rightarrow \underline{R} \rightarrow \underline{S}$. The data from the statistical tests refer to the value of $\underline{K}$, i.e. the proposition $\underline{S}$. The inference which is usually, but erroneously drawn, is that the evidence supporting $\underline{S}$ in turn supports $\underline{R}$ which in turn serves as indirect support for the theory embodied in $\underline{Q}$. But, as has been noted before, the only way in which evidence for $\underline{S}$ can be used to corroborate $\underline{R}$ is if we have independent evidence supporting either $\underline{Q}$ or $\underline{R}$. In the above formulation $\underline{Q}$ and hence $\underline{R}$ contain a number of unobservable initial conditions. Since the presence of

these conditions precludes the possibility of directly testing the propositions in $\underline{Q}$ and $\underline{R}$, the evidence supporting $\underline{S}$ cannot be employed to support the proposition $\underline{Q} \rightarrow \underline{R} \rightarrow \underline{S}$.

Clearly, the Cobb-Douglas function $\underline{R}$ can be taken by itself, without any of its usual theoretical underpinnings, and put forward as an observed statistical regularity. That is to say, $\underline{R}$ can be proposed as an hypothesis standing by itself and can be fitted to the appropriate data. Under these conditions the general proposition would now be restricted to $\underline{R} \rightarrow \underline{S}$. To corroborate this proposition one would still need two sets of data, one supporting $\underline{R}$ and one supporting $\underline{S}$. But, as $\underline{R}$ has been detached from its unobservable antecedents, such support is no longer theoretically impossible.

The same comments, of course, apply to the Simon-Levy proposition which can be represented by $\underline{R}' \rightarrow \underline{S}$ where $\underline{R}'$ is the hypothesis given by (9.4). Since the same data support the consequent $\underline{S}$ in both cases, an independent set of data would have to be found which supported $\underline{R}'$ before the proposition $\underline{R}' \rightarrow \underline{S}$ could be said to be empirically confirmed.

Nothing, so far, has been mentioned about how one might subject propositions like $\underline{R} \rightarrow \underline{S}$ or $\underline{R}' \rightarrow \underline{S}$ to the requisite empirical tests. This topic will be discussed next. At present my concern is to point out the fact that as long as econometric hypotheses are considered as consequents of general economic theory, such as $\underline{Q} \rightarrow \underline{R} \rightarrow \underline{S}$, then the presence of equilibrium conditions will preclude the possibility of ever subjecting these hypotheses to empirical test. On the other hand, if econometric hypotheses are viewed as standing by themselves for empirical appraisal then, whatever their actual or theoretical origins, it is at least possible to observe the relevant initial conditions preparatory

to conducting empirical tests. In this respect an hypothesis' origins are irrelevant, and the important questions which remain--can this hypothesis be tested? is this hypothesis empirically true?--can now be investigated.

2. Testing Econometric Hypotheses

In order to submit any hypothesis to a process of refutation by empirical test it is necessary to have a procedure which will identify those data that will disconfirm the hypothesis. The process of testing is, in fact, a process of searching for negative results. Unless the testing procedure delimits those data which are to be considered instances of disconfirmation, the testing process cannot be carried out. The growth of a body of scientific theory is predicated upon the detection of erroneous hypotheses. And unless it is possible to identify the disconfirming instances it is not possible to detect the errors. Consequently, when examining a testing procedure the principal item to look for is the process by which an hypothesis is rejected. If it is not possible to reject certain hypotheses then it is not possible to decide whether they are empirically true or false. Thus, in this examination of econometric hypotheses the object is to investigate the procedures, if any, by which this class of propositions can be rejected by empirical test.

To clarify the difficulties which surround the testing of econometric hypothesis it is easiest to begin by examining the general problem of testing statistical hypotheses. As noted in Chapter 7 a statistical hypothesis is a statement about the probability density function of a random variable. The density function of a random variable refers to the assumed or actual density function which characterizes the population of which the variable is a member.

When a sample of observations is taken it is drawn from this population. And if the test is concerned with determining the actual density function describing this population, it is the sample data that are used to perform the test.

In the simple example employed to illustrate the principle of maximum likelihood there was an urn which contained a population of red and white balls. We happen to know that this population is characterized by the binomial distribution. But if we were presented with this urn without knowing the distribution of balls inside it we could determine its density function by a number of methods each of which relies on our ability to draw repeated samples from the urn. If the population of the urn were large enough the sampling could be conducted without replacement. For the size of the population would prevent the withdrawal of the individual samples from distorting its actual density function.

Suppose, for example, that we have an urn which is filled with a large number of colored balls. In all, there are four different colors red, green, white and black. The population of this urn is generated by a particular biological process about which we have a theory. One of the hypotheses of this theory concerns the frequency of occurrence of the phenomena we have called colored balls. Under appropriate conditions this hypothesis states that the different colors are present in the population in the ratios $9_{\text{red}} : 3_{\text{green}} : 3_{\text{white}} : 1_{\text{black}}$. We are unable to observe and count the actual frequencies in the population, but the population is large enough so that we can sample without replacement.

In order to test this hypothesis we draw a number of samples from the population and record the total number of occurrences of each of the colored balls.

From the theory, the hypothesis states that the probabilities of occurrence are $p_{\text{red}}=9/16$, $p_{\text{green}}=3/16$, $p_{\text{white}}=3/16$, $p_{\text{black}}=1/16$. The sample data provides the observed frequencies of their occurrence which can be directly compared to the theoretical by multiplying the theoretical probabilities by the total sample size as shown below.

	Red	Green	White	Black
Observed Frequencies	240	96	72	26
Theoretical Frequencies	244	81	81	27

Total Sample Size = 343 Observations

To test the null hypothesis we employ the Chi-Square statistic. Computing its value from these data the result is $\chi^2 = 3.9$. In this case there are three degrees of freedom, and if the rejection region is set at 5% the critical value of the statistic is given by $\chi^2 = 7.8$. Since the computed is less than the critical value we do not reject the null hypothesis. Clearly, if the computed had exceeded the critical value we would have rejected the hypothesis.

The process of accepting or rejecting an hypothesis by comparing a computed to a critical value is based upon the notion of a confidence interval and the critical region from which these intervals are built. A critical region of 5% implies that under the assumption the null hypothesis is true the computed value will lie within this region 5% of the time. Thus, a critical region like a confidence interval rests on the tacit assumption that the experiment can be repeated a large number of times. To repeat an experiment the population from which the samples are drawn must remain unchanged. In actual practice minor changes can take place in a population without significantly affecting the testing procedure. But as long as the major factors are known and

controllable, repeated sampling can be employed to test hypotheses about the nature of the underlying population.

While this simple example and the ancillary comments are undoubtedly obvious to the reader their full force does not seem to be appreciated by practicing econometricians. For econometric hypotheses differ in several important respects from the one employed above. To begin with econometric hypotheses are stated in terms of endogenous and exogenous variables, and parameters. In the above example the theory about the biological process asserted the existence of a particular ratio of probabilities of occurrence. These probabilities are the counterparts to the parameters in an econometric hypothesis. Hence, one would expect to be able to determine the "actual" value of the econometric parameters in the same way that the actual values of the probabilities were ascertained. To determine the "true"^{8/} parameter values of a population density, it is necessary to have a stable population--namely, one from which repeated samples can be drawn. Unfortunately, the population of economic variables is quite unlike that of the urn. Not only does one not have any assurance that the population remains the same from one time period to the next, but one is also unable to control the principal factors effecting such changes. Consequently, samples drawn at different periods of time cannot be shown to come from the same theoretical urn and little can be done to alleviate this problem.

Another significant difference between the exemplar and an econometric hypothesis occurs in the manner in which the sample data are normally employed.

^{8/}That we can never know the exact values of the population's parameters, goes without saying. But, as long as the population density remains unchanged we can come as close to the true value as we have the time and the patience to sample.

In econometrics the sample data are used to estimate the parameter values. By calculating the standard errors of the estimates, confidence intervals can be constructed to adjudge the estimators' reliability. The notion of a confidence interval, however, depends on the assumption that the same population can be repeatedly sampled. Since this condition does not strictly apply to econometric investigations this measure of an estimate's reliability is only relevant in a loose and qualitative way. When dealing with the urn it is possible to test the observed sample estimates against the known (and knowable^{9/}) population parameters. To do so the null hypothesis is constructed from the actual population values and the sample estimates are employed to reject or not reject H_0 . For an econometric hypothesis, however, what is the null hypothesis? Since the actual parameter values are not ascertainable against what set of values is it possible to test?

An immediate answer is to employ the null hypothesis where the parameter values are set equal to zero. For example, if one collected data to estimate the values of the parameters in (9.1) the relevant null hypothesis would be:

$$H_0: \alpha = 0 \quad \beta = 0$$

As noted in Chapter 7, from the standard errors of the sample estimates $\hat{\alpha}$ and $\hat{\beta}$ one can set up a confidence interval for $\hat{\alpha}$ and $\hat{\beta}$. If these intervals include $\hat{\alpha} = 0$ and $\hat{\beta} = 0$, then according to normal procedure one is not able to reject the null hypothesis. However, since it is not possible to conduct repeated samples there does not appear to be any reasons why this test is

^{9/} The population's parameters are knowable in the sense that it is in principle possible to ascertain their values to whatever degree of accuracy one cares to chose.

appropriate in the first place. Even though the sample size may be increased to enhance the statistical significance of the results, the inability to repeatedly sample from the same population reduces the rigor of the test.

In order to circumvent this obstacle one would have to know (have a testable and tested theory of) the process by which the elements of the population are generated. In the case of the urn a biological theory containing both testable and tested hypotheses accounts for the process by which the relevant ratios are derived. While this theory may eventually be replaced by another, at the present moment it asserts that the generating process is of a certain type with specific, identifiable and testable characteristics. Consequently, even though various factors may affect this process from time to time, the effect these influences have on the characteristics of the population can be evaluated from a knowledge of the generating process. In econometrics the process or processes by which the members of a specific population are generated are unknown--i.e. econometrics does not contain tested theories of these processes. Manifestly, it is this lack of knowledge of the process which prevents one from being able to demonstrate that it is possible to repeatedly sample from the same population.

Summary and Conclusion

At the beginning of this chapter the empirical content of the initial conditions surrounding the error term was examined. As any competent statistician would readily avow, there are a number of statistical tests with which it is possible to determine the presence or absence, in any particular case, of the five basic assumptions. These tests depend, like all statistical tests, on the presumption that the population being sampled from remains unchanged throughout

the sampling process. Further, the method of constructing the null hypothesis and employing sample data to confute or support it can only yield statistically and hence empirically significant results if the null hypothesis concerning the unchanging population is supported by test. To conduct such tests it is necessary either to know the process by which the population is generated or to be able to independently measure the population's relevant characteristics. Since it is not possible to independently measure the "true" values of an econometric population's characteristics, it follows that to employ statistical tests the process by which these populations are generated must be known. To be required to have a knowledge of the process is another way of saying that one needs to have a testable and tested theory of the process. If such a theory is to be stated in econometric (stochastic) terms one will again require some further independent means for checking on the population's stability.

It appears, therefore, that the testing of econometric hypotheses is caught in a moderately vicious circle. In order to ascertain whether the initial conditions of the error term are satisfied a knowledge of the stability of the relevant characteristics of the population is required. Similarly, if one is to adjudge the reliability of parameter estimates, the basis of this measure lies in the assumed ability to repeatedly sample from the same population. Moreover, it is not possible to test the estimated hypothesis against such simple null hypotheses as, $H_0: \alpha = 0, \beta = 0$, unless the stability of the population can be empirically established.

One way of answering this problem would be to appeal to classical economic theory as the theory about the process by which the econometric populations are generated. If this body of theory, or any of its principal hypotheses, could be corroborated by empirical test then these hypotheses could be employed as the

basis from which the independent checks on the population could be carried out. But, to invoke the hypotheses of classical economic theory is to require one to be able to empirically identify the presence or absence of the relevant equilibrium conditions. Since this body of theory does not contain sufficient interpretive rules to allow the initial equilibrium conditions to be empirically investigated, it is not possible to confute these hypotheses by empirical test. To be unable to test these hypotheses is to render them incapable of performing the requisite service. As a result, a knowledge of the relevant processes cannot be acquired by appealing to classical equilibrium theory. To do so is to ensure that the resulting econometric theory is completely untestable.

A second approach to a solution would be to adopt a somewhat pragmatic approach to econometric theory itself. If all econometric hypotheses and models are considered by themselves as statements about the behavior of certain economic variables, then the empirical corroboration of these propositions rests upon two possible grounds. The first, concerns the empirical basis for the statistical tests that are employed. This basis principally requires the population from which the samples are drawn to remain stable throughout the testing process. But econometric theory does not contain the requisite theoretical statements and interpretive rules by which this stability can be empirically ascertained. Consequently, there is no way by which the suppositions entailed in the first approach can be supported by empirical test. If neither the hypothesis nor its sample estimates can be tested on direct statistical grounds, there is only one other possible source of evidential support--namely, the use of forecasts or predictions as the basis for empirical tests. It is toward an examination of this approach to empirical corroboration of econometric hypotheses that the next chapter is devoted.

Chapter 10

Explanation and Prediction in Econometrics

To establish an explanation for the occurrence of an event it is necessary to be able to deduce the phenomena from the conjunction of the theory's hypotheses and the relevant initial conditions. Further, the initial conditions must be empirically true and the theory itself must contain at least one hypothesis that survived empirical tests.

In the previous chapter the empirical content of econometric hypotheses was explored with special attention being paid to their initial conditions. If, as I have argued is the case, the presence or absence of an hypothesis' initial conditions cannot be established by statistical test then one of the basic requirements for a scientific explanation is not satisfied. Since all econometric hypotheses contain error terms and since, without a measure of the population's stability, neither the conditions surrounding the error term nor the reliability of the parameter estimates can be assessed, econometric hypotheses cannot be employed to establish explanations of economic events. Consequently, the question immediately arises as to whether predictions generated by these hypotheses can be employed as a means for subjecting them to test.

For example, consider the simple demand relation used in the previous chapter.

$$y_{(t)} = \alpha + \beta p_{(t)} + u_{(t)} \quad (10.1)$$

One way to test this hypothesis, it might be argued, is to collect a set of data during period $\underline{t}$ on the quantity demanded of a certain commodity, $y_{(t)}$, and on the market price at which these transactions were carried out, $p_{(t)}$. If these data were divided into two lots, one lot could be used to develop the estimates

for $\underline{\alpha}$, $\underline{\beta}$ and $\underline{u}(t)$. From these data the estimated relation would be developed

$$y(t) = \hat{\alpha} + \hat{\beta} p(t) + \hat{u}(t) \quad (10.2)$$

as well as the standard errors of these estimates. The next step would be to employ the second set of data to test the estimated relation (10.2). This test could be conducted in a number of ways, one of which would be as follows: Take relation (10.1) and use the new data to estimate the parameters once again. The result is a new set of estimates which can be represented by

$$y(t) = \hat{\alpha}_1 + \hat{\beta}_1 p(t) + \hat{u}_1(t) \quad (10.3)$$

To test (10.3) against (10.2), adopt the estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{u}(t)$ as the values for the null hypothesis. By employing the standard errors and sample size associated with (10.3) one can, for a specific size of Type I error, develop the required confidence intervals about $\hat{\alpha}_1$, $\hat{\beta}_1$, and $\hat{u}_1(t)$, and determine whether the null hypothesis is to be rejected or not. The procedure is based on the prediction that the relation estimated from the first set of data will hold for the second set as well. Clearly, if the total pool of data is large enough it can be broken up into a number of subsets, and several of these tests can be conducted.

For a test to have empirical significance it must be possible for the data to disconfirm the stated hypothesis. Hence, if these checks on the estimated relation (10.2) are to serve as legitimate tests it must be possible to reject the null hypothesis. But what conditions must be met if it is to be possible to reject the null hypothesis? As in the case of any empirical test on a specific hypothesis it must be possible to empirically identify the presence of the initial conditions. It has already been shown, however, that one cannot establish the initial conditions surrounding the error term as empirically true. Even

though these tests may well produce supporting evidence for the null hypothesis given by (10.2) none of these tests can serve to disconfirm it. Accordingly, this application of (10.2) as a predictive device within period, t , does not provide a method for testing the basic hypothesis given in (10.1).

Since the first approach is not successful, a second attempt at a solution is offered by employing the estimated relation of period t to predict the price and quantity relation to be observed in the following period. That is to say, if the data from period t are employed to generate estimates for (10.1) the estimated hypothesis can then be used to predict or forecast the relation which occurs in period $t+1$. If the data from $t+1$ support the estimate based on period t , this result is said to confirm the basic hypothesis given by (10.1). In order to examine the merits of this claim it is necessary to review briefly the conditions under which a prediction can be employed to test a particular hypothesis.

1. Prediction and Empirical Tests

The structure of a scientific prediction is similar to that of an explanation. In the case of a deterministic (non-statistical) hypothesis a prediction can be used as a test of the hypothesis if the same three conditions are met:^{1/} First, the predicted event must belong to the class of observables. Second, the theory or hypothesis must be open to rejection by empirical test. Third, for the second condition to be satisfied the relevant initial conditions

^{1/}See Chapter 2, pp. 26-28.

must be empirically true. Hence, if an hypothesis is represented by $\underline{R}$, and the deduced or predicted relation is represented by $\underline{S}$, the structure of the prediction is given by the proposition $\underline{R} \rightarrow \underline{S}$. As has already been noted, evidence supporting $\underline{S}$ can be counted as confirming $\underline{R} \rightarrow \underline{S}$ if there is independent evidence supporting $\underline{R}$. Part of such evidence is data demonstrating the presence of the relevant initial conditions. If $\underline{R}$ is a consequence of the theory $\underline{Q}$, then some of the initial conditions may also be a consequent of $\underline{Q}$. But if $\underline{R}$ is considered by itself, the specification of $\underline{R}$ implies that certain initial conditions be satisfied. As long as these initial conditions can be subjected to empirical test, then the proposition $\underline{R} \rightarrow \underline{S}$ can indeed be confirmed or disconfirmed by the evidence pertaining to $\underline{S}$.

However, when dealing with a statistical hypothesis the problem becomes slightly more complex. Consider, for example, a statistical hypothesis which relates the occurrence of two properties of certain events by a specific probability. Suppose the probability that an occurrence of $\underline{M}(x_i)$ will also be an occurrence of $\underline{N}(x_i)$ is given by $\underline{r}$ (where $\underline{r}$ is the long-run frequency of all $\underline{M}(x_i)$'s being $\underline{N}(x_i)$'s), then the hypothesis can be stated in the standard form: for all x_i , $p[\underline{M}(x_i), \underline{N}(x_i)] = \underline{r}$.

Given such an hypothesis how does one subject it to test? To begin with it is perfectly clear that, even though the relevant initial conditions can be shown to be satisfied and that $\underline{M}$ and $\underline{N}$ are observable attributes of x_i , one contrary instance is not sufficient to reject the hypothesis. For whatever the value of $\underline{r}$, say $\underline{r} = .95$, there are bound to be instances where an occurrence of $\underline{M}(x_i)$ is not also an occurrence of $\underline{N}(x_i)$. Indeed, one would expect this to happen $\underline{1-r}\%$ of the time. Thus, to disconfirm such an hypothesis it must be

possible to estimate from the test data whether the observed frequency is significantly different from the hypothesized value, $\underline{r}$.^{2/}

To determine the statistical significance of the observed frequency a null hypothesis is set up which employs the hypothesized $\underline{r}$. By using the appropriate statistical test and accepting a specific Type I error one can then ascertain whether the null hypothesis is to be rejected or not. The basic assumption underlying this procedure is that this test can be repeated a large number of times. If a Type I error of 5% is accepted one is essentially stating that one expects to reject the null hypothesis 5% of the time when it is in fact correct. In the same respect, the Type II error states the percentage of the time one expects to accept the null hypothesis (i.e. not reject it) when it is in fact false. None of these percentages have any statistical meaning unless it is possible to repeat the test. To repeat the test it must be possible to repeatedly sample from the same population. And, as noted above, this condition requires the services of some device or theory which enables one to measure or account for the stability of the population. Therefore, to be able to employ a prediction as a test of a statistical hypothesis it must be possible to repeat the test. If this condition cannot be met then the result of a test cannot be employed to confirm or disconfirm the hypothesis.

2. Econometric Forecasts as Predictive Tests

To help determine whether forecasts can be employed as a means for subjecting

^{2/}A detailed analysis of the problems posed by the use of statistical hypotheses in explanation and prediction can be found in: C.B. Hempel, "Deductive-Nomological vs. Statistical Explanation," in H. Feigl, et al., (eds), Minnesota Studies in the Philosophy of Science, Vol. III, University of Minnesota Press, Minneapolis, 1962, pp. 98-169.

econometric hypotheses to test, a brief examination of the forecasting procedure is in order. To generate a forecasting relation one begins with the basic hypothesis and estimates the relevant parameters from a sample of data from, say, period $\underline{t}$. The result, with the error term deleted, is the forecast relation:

$$y_{(t)} = \hat{\alpha} + \hat{\beta} p_{(t)} \quad (10.4)$$

The error term is ignored because even though a particular set of data may produce an error estimate, such as $\hat{u}_{(t)}$ in (10.2) or (10.3), the expected value of $\hat{u}_{(t)}$ is assumed to be zero. Once (10.4) is developed for forecast for period $\underline{t+1}$, $y^F_{(t)}$, is generated by substituting into (10.4) the observed value for $p_{(t+1)}$. This method of producing forecasts is employed whether the theory is given by the simple relation (10.1) or is represented by a many equation model. Consequently, the question that is being raised can be simply stated as follows: Under what conditions can the forecast value $y^F_{(t+1)}$ be used as an empirical test of the original hypothesis $y_{(t)} = \alpha + \beta p_{(t)} + u_{(t)}$?

From the discussion in the previous section it is clear that a single negative instance is not sufficient to disconfirm the hypothesis. For even though the error term has ostensible disappeared, (10.4) is still a statistical relation and can only be confuted by appropriate statistical test. Hence, to compute $y^F_{(t+1)}$ and then to compare it to the observed value of y in period $\underline{t+1}$ is not sufficient by itself. To adjudge the empirical significance of this comparison it is necessary to examine the statistical test by which these two values can be related.

The standard econometric procedure is to construct an interval about

$\bar{y}^F(t+1)$ and then examine whether the observed value falls within this interval.^{3/}

The interval represents the null hypothesis and is given by:

$$\bar{y}^F(t+1) \pm K \hat{\sigma}_{\bar{y}^F(t+1)} \quad (10.5)$$

where the value of K depends on the size of the sample and the percentage of values of $\bar{y}(t+1)$ to be included in the interval.

The interval (10.5) is a tolerance interval not a confidence interval. And the difference between the two is readily apparent once one examines them a little more closely. A confidence limit or interval for the sample mean of a normal population is of the form, $\bar{x} \pm KS$, where $\bar{x}$ is the sample mean, K the number of standard deviations (defined by the size of the Type I error), and S is the sample standard deviation (error). Moreover, confidence limits are computed in such a way that they will include the actual mean of the population distribution in a fraction γ , where $\gamma = 1 - \text{Type I error}$, of the total set of samples which are gathered. Tolerance intervals, on the other hand, while of the same form, i.e. $\bar{x} \pm KS$, are computed so that within the fraction γ of the samples they will include at least a fraction P of the items in the distribution.

For example, for a normal population the confidence interval $\mu \pm 1.96\sigma$ includes 95% of the population. The tolerance interval $\bar{x} \pm KS$, for the same population, however, is computed so that in a large series of samples the fraction γ of the intervals will include at least P of the population. If the population remains unchanged and if one is free to gather numerous samples, then γ becomes the measure of the degree of confidence with which the tolerance interval can be said to include at least P of the population.

^{3/}The procedure is described more fully in Chapter 8, sec. 4.

Returning to the econometric relation it is clear that since $y_{(t)}$ and $p_{(t)}$ are random variables it follows that $y^F_{(t+1)}$ is also a random variable. If $u_{(t)}$ is normally distributed then $y_{(t)}$ and hence $y^F_{(t+1)}$ are normally distributed. Further, since $y^F_{(t+1)}$ is a random variable it also follows that the estimate of the sample standard deviation (error), $\frac{\hat{\sigma}}{y^F_{(t+1)}}$, is a random variable. Consequently, for each specific sample the variables $y^F_{(t+1)}$ and $\frac{\hat{\sigma}}{y^F_{(t+1)}}$ will have different values. Accordingly, the tolerance interval will vary from sample to sample. While this result does not constitute a special problem it must be possible to repeatedly sample from the sample population before the tolerance interval can be employed as the null hypothesis. For, if there is complete freedom in sampling the value of K (which depends on the sample size) can be chosen so that the probability, γ , that the interval will include at least p of the population is as close to 1 as is desired.

The problem in econometrics is aside from specific sampling difficulties, whether the population remains unchanged. If one is dealing with stable populations then the reliability of the forecast can be assessed by the tolerance interval. But, once it is not possible to adjudge the population's stability, other than by finding out that the forecasted value is in error, then it is no longer possible to employ tolerance intervals to test the forecast relation.

In the previous chapter the problem of testing the initial conditions was examined. Here it was demonstrated that the initial conditions could be empirically checked if and only if the population from which the samples were drawn remained unchanged throughout the sampling process. Further inspection of econometric hypothesis revealed the fact that they do not contain sufficient interpretive rules to permit independent checks on the population's stability

to be made. As a result, it was concluded that it was not possible to empirically test for the presence or absence of the error term's initial conditions. If econometric hypotheses do not contain sufficient interpretive rules to permit the determination of a population's stability this absence of rules also precludes the possibility of submitting the forecasting relation to test. For consider once again the nature of the predictive test. From an estimated equation, such as (10.4), one derives the forecast $\underline{y}^F(t+1)$. By sampling in period $(t+1)$ one computes the observed value of $\underline{y}(t+1)$. Either $\underline{y}(t+1)$ falls within the tolerance interval about $\underline{y}^F(t+1)$ or it does not. But in either event this result cannot serve to confirm or disconfirm (10.4). In order for the evidence to support or deny the forecast relation it must be demonstrated that the population from which the samples were drawn has remained unchanged. Manifestly, such a demonstration is not possible within the bounds of econometrics.

The same argument applies to the case when $\underline{y}(t+1)$ falls outside of the tolerance interval. For without independent knowledge of the behavior of the population this result may well have occurred because of some shift in the population. Such changes are customarily referred to as structural shifts. In the event that they occur the econometrician has no recourse except to re-estimate his original relation and hope there will be no more shifts the next time a forecast is made. While such procedures may well serve the pragmatic test of "usefulness" they are not sufficient to permit the establishment of testable econometric theories or hypotheses.

3. Ceteris Paribus and Population Stability

Part of the inability to test econometric hypotheses stems from the implicit use of the ceteris paribus clause. For, if while the sampling and computations

necessary to check on the forecasted values are being carried out everything else remains unchanged then the testing of econometric hypotheses becomes a practical possibility. Since econometrics does not contain a sufficient set of rules to permit the items in the ceteris paribus clause to be checked one alternative is to examine the effects of ignoring these items on the behavior of a dynamic system. That is to say, if it can be shown that the position of the hypothesized system in period $(t+i)$ will be approximately the same whether the items in the ceteris paribus clause are taken into account or not then, according to this view, there are grounds for ignoring the factors covered by the clause. This approach is based upon two related theorems^{4/} which deal with the problems of analyzing and testing the empirical validity of a dynamic system whose variables are in turn related to other variables not explicitly included in the system--these variables are either assumed to be constant or are merely placed in the general repository, the ceteris paribus clause.^{5/}

In order to explicate this position it is necessary to introduce the concepts of completely decomposable and decomposable systems. A set of relations, theory, or system is completely decomposable if the values of their variables are only a function of past or present values of the same set of variables. Thus, a closed

^{4/} H.A. Simon and A. Ando, "Aggregation of Variables in Dynamic Systems," Econometrica, Vol. 29, April 1961, pp. 111-138; and A. Ando and F.M. Fisher, "Near-Decomposability, Partition and Aggregation, and the Relevance of Stability Discussions," International Economic Review, Vol. IV, January 1963, pp. 53-67. (Both of these articles are reprinted in A. Ando, F.M. Fisher, and H.A. Simon, Essays on the Structure of Social Science Models, M.I.T. Press, Cambridge, 1963, pp. 64-91 and pp. 92-106 respectively.)

^{5/} A more detailed discussion of the problem and the results so far obtained to which this section is to be found in, F.M. Fisher and A. Ando, "Two Theorems on Ceteris Paribus in the Analysis of Dynamic Systems," American Political Science Review, Vol. 56, March 1962, pp. 108-113. (Reprinted in Ando, Fisher, Simon, op. cit., pp. 107-112.)

system (one in which all factors are accounted for) is a completely decomposable system. A decomposable system, on the other hand, is one in which outside factors affect the behavior of the system while the system cannot affect the values of these outside variables. In other words, if a given system is influenced by certain, specified external factors (e.g. exogenous variables) but the system itself is unable to affect the values of these variables, and if these external factors are the only external influences on the system, then the system is called decomposable.

Since neither completely decomposable nor just plain decomposable systems occur with any apparent frequency in economics the two theorems deal with situations which are close approximations to these ideal states. The first, the Simon-Ando theorem, is concerned with systems that are approximately (nearly) completely decomposable--the variables within the system do depend on past values of some outside variables but the dependencies are quite weak in comparison to the internal relations. Given such a system (theory) the theorem asserts that either in the long- or short-run if these external dependencies are ignored the results obtained by treating the system as completely decomposable will be approximately valid. Accordingly, if the external dependencies are relatively weak, the relative behavior of the system treated in isolation will not differ substantially from the behavior the system would have produced if the external factors were taken into account.

The second, Ando-Fisher theorem, states that a similar result holds for systems which are approximately decomposable. That is to say, for systems in which the external dependencies are all one-way and are too large to ignore, the relative behavior both in the long- and the short-run of the system treated

in isolation will not differ substantially from that which the system would have generated if these dependencies were incorporated into the system. "Thus the economist who takes tastes and technology as influencing but uninfluenced by economic variables will find--provided that such an assumption is nearly correct--that his results will be approximately valid in all respects in the short run and that even in the long run, when the full effects of feedbacks in the causal structure are felt, the internal, relative behavior of the variables he studies will be approximately the same."^{6/}

To agree, however, that a particular econometric theory can be said to be nearly completely decomposable or only nearly decomposable does not alter the basic obstacles which confront any attempt to submit the theory to empirical test. In either case the temporary exclusion of the contents of the ceteris paribus clause does not bring one any closer to being able to measure the stability of the underlying population. What these theorems permit one to do is to break down some rather large system, such as a whole economy, into a number of relatively independent parts. This is an important step in the analysis of complex systems. But such an advance does not obviate the necessity for being able to independently assess the characteristics of the underlying population.^{7/} Even though these two theorems permit the econometric theory to exclude those variables in the ceteris paribus clause the statistical form of the theory remains

^{6/}A. Ando and F.M. Fisher, in A. Ando, F.M. Fisher and H.A. Simon, op. cit., p. 109.

^{7/}For further discussion see: F.M. Fisher, A Priori Information and Time Series Analysis: Essays in Economic Theory and Measurement, North Holland Publishing Company, Amsterdam, 1962.

unchanged. Indeed, if econometric theories are to be corroborable by empirical test the basic requirements for the statistical tests must be empirically satisfied. Consequently, until it can be demonstrated that it is possible to repeatedly sample from the sample population the statistical tests are devoid of empirical significance.

4. Micro-Analysis and Statistical Tests

The principal obstacles confronting the testing of econometric hypotheses also appear to encompass the detailed, micro analytic investigations currently being carried out under the aegis of the Social Systems Research Institute.^{8/} The long-range objective of this research is to build a realistic dynamic model of the United States Economy. To construct such a model the economy is represented as consisting of a number of major components each of which is composed of a large number of microcomponents. Accordingly, the behavior of the economy is hypothesized to result in part from the interactions of the microcomponents within each of the major segments, where the components "include markets, goods, and microcomponents such as individuals and families imbedded within regional household sectors and firms imbedded within industries."^{9/}

One of the basic types of components is a "decision unit" and these interact with each other by means of components representing "markets." To complete the interaction between decision units and markets a third type of component is employed which is designated "goods." This last type of component includes all

^{8/} For a recent statement of the approach, objectives, and general conceptual framework of this research see: Guy H. Orcutt, "Microanalytic Models of the United States Economy: Need and Development," American Economic Review, Vol. LII, May 1962, pp. 229-240.

^{9/} Guy H. Orcutt, op. cit., p. 231.

items which are exchanged, produced or consumed by the decision units. Components are described by their input, output and status variables. If the inputs are viewed for the moment as exogenous variables then the behavior of the component is determined by the relations that link the exogenous variables to the status variables which in turn generate the values of the output variables. Each decision unit has a number of outputs some of which become the inputs to the market components which in turn distribute these items, as outputs of the markets, on to other decision units. Since the output of one component frequently is the input of another the principal function of these inputs (output) is to update the values of the appropriate status variable and by means of the behavioral relations generate new values for the outputs.

If the interactions between the major components is slight these parts of the model could be viewed as nearly completely decomposable. As a result, they could be analyzed and investigated independently of the rest of the system. If within each of the major components, the interactions between decision units, markets, and goods are such that the causal relations^{10/} can be specified, then it may well occur that some of these components can be described as nearly decomposable. Under this condition such components may also be treated as relatively independent units.

In order to examine the problems posed by subjecting such a model to empirical test it is perhaps easiest to first consider one of the nearly

^{10/} The meaning of this use of "causality" is presented and employed in: H.A. Simon, "Causal Ordering and Identifiability," W.C. Hood and T.C. Koopmans, (eds), Studies in Econometric Method, Cowles Commission Monograph No. 14, Chapter 3, Wiley, 1953. (Reprinted in H.A. Simon, Models of Man, Chapter 1, Wiley, 1957, and in A. Ando, F.M. Fisher, and H.A. Simon, Chapter 1, op. cit.)

decomposable microcomponents. Whether a particular model contains nearly decomposable units or not is immaterial to the problem at hand. For if it is not possible to confute a relatively isolated component it will not be possible to subject larger portions, up to and including the entire model, to a process of refutation by empirical test.

Consider, then, an approximately or nearly decomposable microcomponent. This component has certain variables which function as inputs, some which are categorized as outputs, as well as its internal status variables.^{11/} The relations which link the inputs to the status variables to the outputs are stochastic. Thus, once a set of inputs are specified these relations specify the probabilities of the occurrence of the outputs. Since the relations are stochastic the problems of submitting them to empirical test are the same as are discussed above--to wit, it must be possible to empirically identify the relevant initial conditions as well as those conditions surrounding the statistical tests. Unless these constraints are satisfied the tests are empirically meaningless.

An example of such a model is provided by recent investigations of some aspects of consumer behavior.^{12/} While the model does not represent the finished product of this research and is still undergoing modification it will aid the discussion to have a specific case to examine. The model is concerned with the

^{11/} The component will be nearly decomposable if most of the inputs for period t are the outputs of period t-1.

^{12/} A.S. Goldberger and M.L. Lee, "Toward A Microanalytic Model of the Household Sector," American Economic Review, Vol. LII, May 1962, pp. 241-251.

demographic and economic behavior of household units and is based upon data collected by the Surveys of Consumer Finance of the Survey Research Center. The relations (operating characteristics) are stochastic and contain the probabilities of the occurrence of the relevant outputs. These probabilities as well as the parameter values are estimated from the survey data and are tested in the usual way for statistical significance. In the model reported in this study there are 23 dependent variables: "total income, home ownership (probability and value of owned home), monthly rent, expenditures on additions and repairs to house (probability and amount, car ownership (probability), new car ownership (probability), multiple car ownership (probability), purchase of car in prior year (probability, total price, and net outlay), purchase of new car in prior year (probability), purchase of household durables in prior year (probability and amount), mortgage indebtedness (probability, amount, and monthly payment), non-car installment indebtedness (probability, amount, and monthly payment), and debt incurred in connection with purchase of car in prior year (probability and amount)."^{13/} The value of each variable, as mentioned above, is estimated by least squares regression from the survey data. And the results are reported along with a notation as to their statistical significance.

While the sample size (3,000 units for each year) and the number of variables considered are considerably larger and the level of detail conspicuously more microscopic than most econometric studies,^{14/} the obstacles to submitting this model to empirical test have not been altered or overcome. This is not to say

^{13/} A.S. Goldberger and M.L. Lee, op. cit., p. 244.

^{14/} The earliest and most complete microanalytic model is presented in considerable detail in: G.H. Orcutt, M. Greenberger, J. Korbel, and A.M. Rivlin, Microanalysis of Socioeconomic Systems: A Simulation Study, Harper and Row Company, New York, 1961.

that if 27 variables were considered instead of 23 the results would not differ from the above. Nor am I suggesting that it is not possible to discard certain variables when it turns out that they have non-significant coefficients. For the difficulty lies not in our ability to regress or manipulate variables and data. The problem is whether such models can ever be disconfirmed by empirical test.

As has been noted above the answer to this question is to be found within the statistical tests that are employed. All statistical tests require specific initial conditions to be satisfied. And all statistical tests assume that repeated samples can be drawn from the same population. In the case where controlled experiments are performed, statistically significant differences between two populations are determined by maintaining a "control" group as well as a "test" group throughout the experiment. In the case of the household model, however, how is one to determine the stability of the underlying population? The data are drawn from sample surveys which are in turn based on detailed interviews of household behavior. If a relation is estimated from the data of one year and then employed to forecast some variable values in the next year how are negative results to be interpreted? Clearly, a shift in the population could account for such a result just as readily as an error in the initial specification of the relation. Further, if the model cannot be considered to be approximately decomposable the negative result could perhaps be attributed to a shift in one of the exogenous factors. If the process by which households made the decisions relevant to the above model were understood, and if it were also possible to test for the constancy of this process, then it would begin to be possible to determine the stability of the population from which the samples are drawn. But, until the stability of the population can be independently measured it is not possible to

disconfirm either this model or any similar microanalytic model by the application of statistical tests.

Summary and Conclusions

At the beginning of the chapter it was suggested that it might be possible to test econometric hypotheses by employing them to generate forecasts or predictions for forthcoming periods. If the forecasts could be compared with the actual outcome it was hoped that this comparison would serve as the empirical basis for the test. For such a test to have empirical significance, however, it must be possible to disconfirm the hypothesis from which the forecast was deduced. With econometric relations this implies that it must be possible to reject the null hypothesis. But to reject the null hypothesis two conditions must be satisfied. The first, which is derived from the standard requirements of an empirical test, requires that the initial conditions surrounding the error term be empirically true. The second, which follows from the basic nature of all statistical tests, requires that it must be possible to draw repeated samples from the same population. Only if a test can be repeated under the same relevant conditions does it make statistical sense to employ confidence intervals and other measures to determine the rejection or non-rejection of the null hypothesis.

In a particular test a specific hypothesis the forecasted value is generated and compared to the observed outcome. For the comparison of these values to reject the hypothesis it must be possible to exclude all other sources of error. If the initial conditions are satisfied and if the population from which the sample values are drawn is the same in both cases, and if it is possible to repeat the test any number of times, then a negative result can serve to disconfirm the

hypothesis. Once the characteristics of a population are empirically determinable then it is possible to test for the presence of the initial conditions as well as for the stability throughout the testing process of the underlying population. Even though the testing of forecasts requires the use of tolerance intervals, if the above factors are empirically determinable, econometric hypotheses can be submitted to test.

However, econometric hypotheses are concerned with the characteristics of certain populations. Indeed, they are no more than hypothesized relations among selected characteristics of certain underlying populations. Further, these hypotheses are not concerned with the processes by which the populations are generated. As a result, econometric hypotheses do not contain the requisite interpretive rules for ascertaining the true values of the population's parameters. Since a shift in the population (structural shift) cannot be determined prior to and independently of a specific test of a particular hypothesis it is not possible to use the results of a test to disconfirm the hypothesis.

The same conclusion holds even for the cases of nearly completely decomposable or nearly decomposable models. For, although in these cases the sources of error are greatly reduced--one is now entitled to ignore the factors in the ceteris paribus clause--it is nevertheless still not possible to disconfirm the model or hypothesis. To do so requires a knowledge of the underlying population and the process by which it changes over time. Consequently, even when dealing with completely decomposable models econometrics does not provide the requisite interpretive rules. Similarly, in microanalytic models where the hypotheses may relate to individual units such as households and the sample sizes are noticeably increased, the fundamental obstacles to empirical testing still have not been removed.

If we cannot disconfirm a hypothesis it is not possible to employ it to establish a scientific explanation or prediction of an economic event. Since we cannot disconfirm econometric hypotheses it follows that we, as economists, are completely unable to explain or predict the occurrence of economic events. While this conclusion may neither surprise nor upset some economists it does reflect rather strikingly upon the state of economics as a scientific enterprise. For, since we are unable to submit our hypotheses to disconfirmation by empirical test, economics cannot be a part of empirical science. It follows, therefore, that none of the conclusions, policies and prescriptions derived from either classical or econometric theory rest upon a testable and tested empirical base.

To me this is an alarming conclusion and one to which a considerable amount of serious thought should be given. If a body of economic knowledge is to be developed it is necessary to have economic hypotheses that can be submitted to empirical test. Since the equilibrium conditions of classical theory appear to be empirically intractable there is little to be gained by searching for a solution in this direction. In econometrics the empirical obstacle is the inability to employ econometric hypotheses to measure directly a population's characteristics. Unless this difficulty can be circumvented econometrics cannot serve as the method for developing testable hypotheses or theories.

In order to rescue econometrics from beyond the pale of empirical science it is clearly necessary to be able to understand the decision processes which govern the behavior of economic units, whether they be individuals, households, markets or firms. Such an understanding would have to imply that a knowledge of the decision processes themselves is sufficient to provide the requisite measures on a population's characteristics and stability. That is to say, for

this understanding to perform the required service a knowledge of the relevant decision processes must lead to empirical measures of a population's stability. Since empirical knowledge is a consequent of testable theory, the acquisition of knowledge about decision processes entails the development of testable theories about such processes. Moreover, once theories are constructed which describe and explain the decision behavior of economic units, these theories may well provide a new basis from which to infer the structure and content of testable econometric hypotheses. While this line of reasoning may strike the reader as a trifle fanciful the remaining chapters are devoted to a detailed exploration of the empirical and practical possibilities of developing a science of economics in this manner.

PART FOUR

Behavioral Foundations
of
Economic Analysis

THE ECONOMIST'S PERSPECTIVE

The fundamental problem facing economists is to acquire a body of empirical knowledge about economic phenomena. Whether such knowledge, once garnered, is used to explain the occurrence of particular economic events or is employed as the basis of public or private policy formulations is not at issue here. For before, knowledge can be used to explain events or to solve problems it must be acquired. And so far the analysis has been solely devoted to the acquisitive process which is a consequence of both classical and econometric theories of economics.

In order to develop a body of knowledge about particular observable events it is first necessary to have one or more empirical hypotheses about the phenomena in question. If these hypotheses survive a number of tests then they themselves constitute the basis for empirical knowledge of this class or these classes of events. Concurrently, such hypotheses also permit one to establish explanations and predictions of the occurrences of the phenomena with which they are concerned.

In the previous two parts of the book the body of economic theory that has been developed both from classical and econometric foundations is examined in some detail. The object of the inquiry is to discover whether either or both of these collections of theory can serve as the basis for a corpus of empirical knowledge of economic phenomena. For a theory to serve this important function at least one of its constituent hypotheses must be confutable by empirical test. But the analysis demonstrates that both the classical and the econometric foundations are such that they preclude the possibility of submitting either type of hypothesis to empirical test. In the former case it is the nonobservable, equilibrium conditions which constitute the empirical stumbling block. While with econometric hypotheses

it is the intractable behavior of the population distribution which prevents statistically significant tests from being performed. As a result, neither class of theory can, in its present state, provide the theoretical basis for a body of empirical knowledge of economic events.

Given such a conclusion it might well be in order to ask whether in fact we, as economists, need testable theories. After all, since neither classical nor econometric theories can be refuted by test, economists have survived remarkably well without the benefits such theories are supposed to provide. To answer this query it is necessary to examine the types of problems to which economic theory in its current condition is applied. For economics is a sufficiently established discipline to suggest that some problems must form reasonably appropriate bases for economic investigation.

One major class of problems to which economic theory has always been applied is the entire range of normative questions and issues. What should be done in such a situation if we wish to maximize (or minimize) some criterion function? What, given certain social and political values, should the best national policy be toward the problems of taxation, unemployment, social welfare, tariffs, money supply, etc. In all of these questions economic theory is used to provide the basis for a "rational" answer. As such the specific set of hypotheses that are employed are used as the instruments ^{1/} with which a solution is generated. Since the desired solution is generated. Since the desired solution is normative, a close correspondence between theory and observed behavior is frequently considered to be unnecessary.^{2/} For the solution is a proposal for what ought to be done.

^{1/} This approach to the justification of a theory is admirably presented by K. R. Popper: "Three Views Concerning Human Knowledge," Conjectures and Refutations, Chapter 3, 1963.

^{2/} See for example: La Von Mises, op. cit.

Indeed, if current behavior is at variance with the proposed then this is just further evidence why a particular solution should be adopted.

In keeping with this prescriptive function of economic theory it is also argued that it is not necessary to completely understand the behavior of economic actors or units for the purpose of developing rules or policies by which they should behave. Accordingly, economists should restrict themselves to finding normative solutions to all economic problems, so that once they are discovered all that need be done is to get the people involved to follow the prescribed policies.^{3/} Indeed, if governments and individuals would only follow the economist's prescriptions and behave according to all the rules, then there would be no need to be concerned over the question whether these theories were testable or not. For in this case the theories would govern economic behavior.

Unfortunately, either governments or individuals are not sufficiently amenable to the economist's persuasion or they are unable to sort their way through the conflicting policies with which they are presented. In either event while economists generate normative theories with their concomitant "rational" solutions the procession of economic events proceeds unexplained. If we were uninterested in the actual economic events themselves this state of affairs would cause little or no concern. But clearly the converse is the case. It is the actual events which affect our lives and create the problems which economic theory is supposed to be able to resolve. And since the theories of economics cannot be refuted by empirical test we as economists and private individuals are left with an unbridged gulf between the economic

^{3/} For an interesting discussion of this issue see: C. W. Churchman, Prediction and Optimal Decision, Prentice-Hall, Englewood Cliffs, N. J., 1961.

events of daily life and policies we are supposed to follow.

While I have no wish to suggest that the formulation of public and private policy is not one of the prime functions of the economist I am willing to argue that this important activity has been approached from the wrong direction. By basing policies on untestable theories there is no possible way of detecting error. Since it sometimes occurs that policy prescriptions are not mutually consistent on what basis is a choice to be made? If economic theory were based on testable theories the answer to such a question would be clear. That is to say, the procedure for arriving at an answer would be given by the theories themselves--submit the conflicting hypotheses to test and see which of them survives or corresponds more closely to the relevant data. Even though such a procedure may not immediately resolve all ambiguities, e.g., the long struggle between the proponents of the wave and corpuscular theories of light, it is the only process by which the conflict can be empirically resolved other than that of resorting to professional fiat.

Further, since science can only progress through the detection of error--progress comes from adopting and creating new theories to account for the inconsistencies and errors in the old--to be deprived of this corrective process is to forever abandon economics to the tyranny of rhetoric and the defense of established positions. Under these conditions criticism on empirical grounds cannot play its vital role.

To avoid this situation, as well as to be in the position where policy formulations can be based on observed behavior, a body of theory must be developed that can be submitted to empirical test. This is not to say that all classical and econometric theory need be immediately scrapped. Rather, the point is that unless and until they can be transposed into empirically testable states, they cannot serve as the basis for a science of economics.

If we are unable to directly employ the results of classical and modern economic thought, how then are we to proceed? And upon what basis and in what direction are we to look for testable hypotheses? While the answers to both these questions may not be immediately obvious, their outlines are explored in the remainder of this chapter.

1. Economic Analysis and the Problem of a River Process ^{4/}

Consider, for a moment, the problem of understanding the behavior of an object which is freely floating on the surface of a river. The river in question empties into a tidal basin where the tides are sufficiently large to affect the river's rate of flow. In fact, at particular periods during the tidal cycle the river, to an observer on its bank, appears to flow in an upstream direction. We, as observers, newly arrived on the scene, are unaware of the tidal properties of the river, and observing the progress of the previously mentioned object feel stimulated to develop a theory to account for its behavior. In keeping with our classical training we immediately perceive the floating object as part of an equilibrium system. We observe a wind as well as the movement of the water and are content to hypothesize that the behavior of the object is determined by the resultant of these two forces.

Gradually, as we stand there congratulating ourselves on our undoubted perspicacity, the object ceases to continue in its "normal" direction. We note that the wind has not perceptibly altered and are moderately puzzled. At this point the object begins to move in the opposite direction and lacking any knowledge about the behavior of the river itself one of us is led to construct the following theory: "Consider the object, if you will pardon the anthropomorphism, as having an overt desire to prolong its stay

^{4/} The example used in this section is directly indebted to: W. Van Orman Quine, From a Logical Point of View, op. cit., Chapter 4

on the river. Since the river must eventually empty into some larger body of water, the object can only prolong its stay if it propels itself against the current. Having only a certain amount of energy at its disposal at any one time it has to decide how to consume this energy to its best advantage. If we posit the existence of a utility surface then it rapidly becomes clear how we are to understand its behavior. For surely, the only rational thing for the object to do is to maximize its utility function subject to its energy constraint."

When questioned on how we might employ this theory to predict the yet unobserved behavior of the object the reply was immediate. "First we must observe the choices made by the object in each period of time. Since it has many alternatives facing it at each instant, the action it takes is clearly that which it prefers. Second, we need a measure of the amount of energy that can be expended per period. Then, since we have the utility surface and the energy constraint we can deduce certain characteristics of its behavior. For example, it will proceed from one equilibrium position to another. And if these equilibria are stable it will remain there until either the wind shifts or some other factor disturbs it. At which point it will strive to return to a new equilibrium position. If you want to deduce further characteristics of its behavior employ the method of comparative statics and determine the directions of the appropriate rates of change."

At this juncture, another member of our company, while agreeing with the theory as stated, proposed a more direct method by which we could predict the future behavior of this object. "Since the object's behavior is in part determined by the river's current and the prevailing wind we should set up a relation including these components and estimate the relevant

parameters by taking a sample of observations. In particular, the dominant direction of the wind should be noted so that the net effect of the wind in this direction can be represented by an error term. Then, if we also measure the distance of the object traverses in a given time interval and introduce an unknown variable to represent the river's speed, we can estimate the value of this unknown by standard econometric methods. Once we have an estimate of the river's speed then, in conjunction with the estimate of the utility surface and the object's desire to maximize same subject to its energy constraint, we can predict for a specific interval of time the future course of the object's progress."

While this example may appear quite fanciful to some readers, let us stop for a minute to examine the position into which we have been led by both of these methods of analysis.

To begin with it should not be forgotten that the problem is to account for the seemingly "odd" behavior of the floating object. The classical approach considers behavior in terms of equilibrium positions. Thus all movement is either toward or away from such equilibria. Concurrently, the equilibria are end points or states of momentary rest, especially if they are stable, and behavior is viewed as a process of proceeding from one end state to the next. To account for the object's progress on the river the classical observer was led to represent its behavior in terms of these equilibria. As a result, the observed behavior was "seen" in these terms and constructs were developed to accommodate this view. Moreover, all behavior exhibited by the object supports this position, since none can confound it.

But what have we learned about the object's behavior? Since the theory cannot be refuted the scientist must reply, "Not a thing! Because, you

have been looking at the wrong sorts of things.^{5/} To focus your attention on equilibrium positions is to put yourself in the same situation as Heracleitus, who long ago complained that he could not bathe in the same river twice because new waters were forever flowing by. What Heracleitus apparently noticed was that he could only bathe at any one instant of time in one stage of the river. At a second moment he could bathe in another river stage but it could not be the same river stage as that in which he had previously been immersed.

"Equilibria are very similar to river stages. No two equilibria can be the same, and each marks another stage in the object's progress over time. To focus on equilibria is to focus on behavior stages. Behavior is a process through time and behavior stages represented by equilibria are at best its momentary parts. But, and here is the point you seem to have missed, to identify the river bathed in the first time with river bathed in once again is precisely what determines the subject matter to be a river process and not a river stage. In other words, by seeing observed behavior in terms of behavior stages you should have been led to study the behavior process not the behavior stages. For to confine your investigations to behavior stages (equilibria) is to forever restrict yourselves to studying things that never remain the same. Science seeks to discover empirical regularities so it is somewhat awkward to be looking for testable relations where none can be found.

"Those of us who labor in the natural sciences are admittedly blessed with the opportunity of conducting controlled experiments when we are confronted with behavior we do not understand. But take your object floating out there

^{5/} For further discussion of this point see: ibid, pp. 65-68.

on the river. Under the circumstances we cannot very well experiment with it so how should we proceed? To understand its behavior we need to know, as you have already correctly pointed out, something about the forces that impinge upon it. In this case we would like to know the forces acting on the object from the water and from the wind. But more than that we need to know something about the behavior processes of the object itself. If, as you have suggested, it is animate, we need to know the process by which it reacts to water and wind. If it is inanimate its behavior is solely the result of the external forces acting upon it. In either event it avails us not to whit to view the observed behavior as anything but the resultant of the interaction of a number of specifiable forces and processes. To hypothesize that the observed behavior is the resultant of an equilibrating process, so that all we can see are the equilibrium stages is, unless we have by other means acquired a knowledge of the processes themselves, to place ourselves in the unenviable position of never being acquainted with more than the equilibrium stages themselves. While we may describe these stages in ever increasing detail, because each one differs from the next, we will never find the empirical regularities we would so much like to find.

"You must forgive me for carrying on at such length and for bringing up the point I should like to mention next.^{6/} But, you see, in the natural sciences we were straightened out on this point quite some time ago. In the very early days of science, Aristotle, if I am correct, thought that all bodies were supposed to want to come to rest. That is to say, all bodies had a natural

^{6/} This part is indebted to the excellent history of science presented by Herbert Butterfield, The Origins of Modern Science, C. Bell and Sons, London, 1957.

place on which or in which they were supposed to want to rest. For example, all heavy terrestrial bodies were believed to have a natural motion towards the center of the universe. And as the center of the universe was believed to be the earth, all such bodies had a natural motion toward the center of the earth. The behavior of objects, then, was viewed as comprising successive stages toward a natural end state or equilibrium stage. Any movement away from a state of rest implied the existence of a motivating force or "mover." As such this was a theory of rest or end states. Consequently, it was motion, not rest, which attracted attention and required an explanation.

"For example, if a body was observed in motion it was assumed that there was a mover actually in contact with it, giving to the object the motion that was observed. Only when the mover ceased to operate could the body come to rest, fall straight to the ground, and arrive at an end state. When you suppose that the object on the river is proceeding from one equilibrium to the next you have also assumed the existence of a mover. In this case you gave a name to the mover and described it as a process of maximization of utility subject to an energy constraint. Clearly your theories about the behavior of consumers, firms and the economic system as a whole are also built in this fashion. But do you really mean to argue that the economy is seeking an end stage or equilibrium point so that once there, if no longer disturbed, it would remain forever in one place?

"Or consider the other side of the same argument. If a body's motion is due solely to the presence of a mover, then, since there is almost always some external resistance to a body's motion, the speed of a body must be proportional to the force being exerted by the mover. If the external resistance is reduced and all other facts remain unchanged, the speed of the body will increase.

Following this line of reasoning to its conclusion we see, as Aristotelians thought, that the speed of bodies in a vacuum (zero resistance) must be infinite. Since they found this an absurd conclusion they rejected the notion of a vacuum, claiming that such a thing could not exist. In a similar manner what happens to an economy, a firm, or a consumer if the frictional forces or resistances to its behavior are reduced? Does the speed of their motion increase in inverse proportion to the resistive forces? Will an economy grow at an ever increasing rate if the frictional forces are reduced? Will the object on the river suddenly start moving at an infinite rate if all frictional forces between it and its environment are eliminated? Does an object freely falling in an approximate vacuum fall with an approximately infinite velocity?

"The answer, in the last case, is obvious for we have certain theories and empirical laws to account for the behavior of freely falling bodies. In particular we have the classical approximation of this behavior--that under normal conditions in vacuo the acceleration of a freely falling body near the earth's surface is given by $\frac{d^2s}{dt^2} = 32$ feet per second. Once again you may well retort that this is a well known empirical regularity which was arrived at by controlled experimentation. But that, I am sorry to say is not really the point. The point is that as long as motion was the phenomenon to be explained, and the end state or equilibrium position the natural place of rest, no one could have observed or discovered this law of behavior.

"It was not until the middle of the seventeenth century that the conceptual framework was sufficiently altered to permit the observation and discovery of such physical regularities. If my memory does not fail, it was Galileo who first realized that motion was not the important thing to explain. Rather it was the change in any particular set of behavior which required the explanation. As you undoubtedly recall, he altered the Aristotelian conception of inertia

so that now a body would either remain at rest or in a uniform motion in any particular direction until some outside force changed that motion. No longer was an object seeking an equilibrium position and once there remaining in a state of rest. Now all objects were either at rest or in a uniform motion and any changes in these motions were the factors to explain.

"The consequences of this view have had, as you are well aware, a profound effect on the course of the physical sciences. And, I am willing to argue, the same would be true of your investigations if only this conception of behavior were taken seriously. For consider once again the object you observed floating in one direction at a fairly steady speed. This motion caused you no surprise as we are all accustomed to seeing objects floating on rivers at moderately constant velocities. But the minute the object's motion changed direction your attention was caught and you began to hypothesize an equilibrium system to account for its behavior. By focusing on equilibria your attention shifted from the original item which caught your notice--the change in direction to a theory which would account for the "mover". Without a mover, an equilibrium approach which seeks to define the end states does not make much sense. Accordingly, the theory you were led to construct was primarily a theory of the mover. And since this kind of mover is not the sort of thing you ought to be looking for, is it any wonder that the result of such an inquiry is an untestable theory of the object's behavior?

"While you may still object to this analysis of your theoretical procedures consider the further difficulties you are led to by the econometric approach. Once again the econometrician is inclined to perceive the object's behavior in terms of movers and equilibria. So much so in fact, that if you inspect a normal econometric model you will find that many of the model's variables come directly from the classical equilibrium conceptual scheme. I refer, of course,

to such variables as the marginal rates of substitution between one variable and another, the marginal propensities to behave in certain ways, and the many other remaining examples of this type. By including these variables within the econometric relations to be estimated you are in effect trying to estimate the various hypothesized characteristics of the mover or movers in question. Since no two equilibrium positions can be the same, no two sets of observations can be guaranteed to be samples from the same population. Not to mention the external forces acting on the object which you either classify as exogenous variables or lump together in the error term.

"In order to develop testable regularities in this fashion you have to know something about the behavior of the population from which you are drawing your samples. For unless you can be sure that the samples come from the sample population it is not possible to submit such relations to test. By focusing your attention on the mover you are unable to learn anything about the population. And once again you are left in the most unsatisfactory position of being unable to empirically test your hypothesized relations.

"For example, consider once again the object floating on the river and the method you proposed for discovering the relation which governs the object's behavior. If I remember correctly, one of you suggested setting up an expression which included the distance travelled by the object during a given interval of time, time itself, the unknown speed of the river, and an error term which expressed the net effect of the wind upon the object. Assuming, at this point, that the object is an animate one you then suggested taking an observation on the distance travelled during the specified interval and using this as your estimate of the object's progress over the assumed constant river speed. As should be apparent by now, this is a procedure for trying to estimate the characteristics of the object's mover. Since the object's behavior

with respect to you standing on the bank did not remain constant over time (it was the change in behavior that caught your attention), how can you possibly hope to estimate the parameters of a regularity from observations that are constantly varying? The only way you can do this is by knowing the process that governs the river's speed. And since, in this case, the tidal effect violates your constancy requirement we are at an apparent dead end.

"Even if the object is inanimate the problem viewed in this fashion is really no different. For how are you going to reconcile the estimate of the river's speed from one instant of time from that derived at some later period? It does not help that these estimates are all derived by maximum likelihood techniques and that each is an unbiased and consistent estimate. If the speed of the river is changing from sample to sample the estimates are all being based on different underlying populations. Accordingly what is really at the heart of the problem, is that you are unable to tell from your estimates alone whether the underlying population has remained the same or not. Indeed, until by some independent check you can be sure of this point it is not mathematically legitimate to employ the appropriate statistical tests.

"Clearly, if you could control the river's progress so that you were assured of its constancy, your methods would produce reliable results. But, unfortunately, you can no more control the river's speed than you can the behavior of the consumers, firms and whole economies which are the more usual subjects of your investigations. Moreover, in these latter cases the behavior patterns are vastly more complex than that of the object floating on the river. If your method of approach cannot lead you to testable relations about this object's behavior how can it possibly succeed when faced with behaviors that are many times more complex?

"The answer, if indeed there is an answer, must lie in an understanding of the processes that govern the behavior under investigation. To understand the process it appears to be necessary to recognize that neither motion nor rest themselves are the prime objects for inquiry. Rather, it is change in motion which should act as the focus of attention. If all objects, whether inanimate or not are perceived as continuing in a state of uniform motion or that of rest until disturbed by an external force, the key to the understanding of the object's behavior lies in discovering the processes which govern the interaction between the object's motion and the disturbing force.

"For example, to understand the behavior of the floating object we need to know both the process that governs its motion as well as the processes by which it interacts with its environment. If the object is inanimate we would all agree that it will continue in a uniform motion downstream until disturbed by wind or contrary river current. Given this conceptual framework, the change in the object's behavior would lead us at once to conjecture some shift in the behavior of the river or the wind. If the wind is observed to be much the same as before we would immediately be led to suspect some shift in the river's motion. Even though we were unaware of the tidal effect we would, without much mental agitation have assumed that some such activity was disturbing the river's flow. Further, if we were sufficiently curious, we could readily corroborate this assumption.

"Given this orderly procession from conjecture to observation, why should we behave differently when the object is no longer inanimate? Admittedly, inanimate objects are easier to handle as one can concentrate almost exclusively on the external forces. But the only additional problem posed by the animate is that we have to understand their internal decision

processes. That is to say, with the animate we have to understand both the internal decision process as well as the processes which govern the interaction between it and its environment. However, by employing the natural sciences as our guide the task should not be as awesome as it may appear. For, under the general hypothesis that all bodies continue in uniform motion until disturbed by an external force, our primary concern is to explain changes in behavior. From an analysis of change we are led to construct hypotheses about the interaction between the environment and the object's decision process. And from an analysis of these processes, if successful, we are led to an explanation of the change.

"Consider, for a moment, how we might proceed by this approach to develop a testable theory to explain the behavior of a consumer, a firm, or an economy. First of all we acknowledge the assumption that we expect the decision behavior of the subject (the consumer, the firm or the economy) to remain unchanged until acted upon by some external force. Such changes in behavior may take place for a variety of reasons, but note that our basic hypothesis leads us to focus on change itself as the event to be explained. To explain this type of event we first need to know a certain amount about the decision process of the subject under consideration. Once we are able to describe and explain such processes we will also be able to identify the external factors that can alter the subject's behavior. For a change in some external factor which is not a part of the subject's decision processes can not very well affect its decision behavior. Consequently, a knowledge of the subject's decision processes will provide us with ability to identify the most likely to the

external influences.^{7/} A knowledge of the object's behavior with respect to its environment suggested the presence of a reverse or tidal current. Similarly, a knowledge of the normal decision behavior of a consumer would suggest the principal factors which would induce him to alter this behavior. While the external events may remain beyond either our own or the consumer's control, a knowledge of his decision processes will lead to hypotheses about the interactive or adaptive process. Once this latter process is sufficiently explored the observed changes can now be explained.

"One further comment and then I will stop. Observe what has been gained by this approach. First, unless your subjects are more recalcitrant than I can imagine, you should begin to discover testable relations governing specific classes of decision behavior. Second, once the first of these has been proposed and tested you will be in a position to employ other observed behavior to test, amend, and generate further empirical hypotheses. At this point you should be in a position to explain the decision behavior of certain individual subjects as well as perhaps that of certain classes of subjects. That is to say, your empirical hypotheses should already have a modest generality.

"Having progressed this far with your empirical knowledge of decision processes it may well be possible to begin employing some of your more established economic techniques. For, if you can specify the principal components of the decision processes of a particular class of economic actors you will have at the same time identified the major external variables

^{7/} Since we are only concerned with decision processes the possible list of external factors such as directly affect the physiology of the organism itself, such as death, crippling disease, etc., are specifically excluded from consideration.

which can affect the behavior of these actors. From a knowledge of the decision process it is possible to detect when this process changes. Since it was the lack of knowledge about such processes upon which your econometric method foundered, it may well be possible to link these two approaches together. In other words, a knowledge of the decision process may provide the measure that is needed to gauge the stability of the econometric population. Once the population's stability can be assured, then econometric relations can be submitted to test. Whether a knowledge of decision processes will provide all the information you need to measure a specific population's stability is a question for you to answer. But the possibility of such a solution should not be overlooked. And in my opinion it would appear to warrant a rather searching examination. At the very least, this whole approach will generate testable hypotheses about economic decision behavior, and with luck it will provide the basis for an empirical science of economic behavior."

2. Classical Analysis and Decision Processes

From the foregoing it is apparent that we, as economists, need to search for testable economic relations which describe the changes in a particular actor's or system's decision behavior or motion. Since physical laws do not state that "A will be followed by B" there is no reason to suggest that economic laws should be framed in this manner either. Rather, like physical laws, we need to develop relations which will tell us how an actor's or a system's behavior is changing at each moment of time, not where the system or actor will be at some future moment.

In this respect, it would appear that the classical economic framework is unsuited to such a task. For the concepts of equilibrium and stability are concepts of states of rest--they describe the points to which the

system may arrive at some future moment, but not how or why the behavior is changing from moment to moment. Since the classical approach is most readily represented by its deductive system the abandonment of these classical concepts entails a departure from this particular deductive system.^{8/}

That this conclusion follows directly from the analysis is readily seen if one considers once again the effect of a conceptual framework upon theoretical development. For it could well be argued that it is the mathematics of the classical system which leads to hypotheses concerning equilibria and which stimulates the discussion of stability conditions, the convexity of sets, and the most suitable axioms for a theory of choice.

If we are to shift our attention from end states to processes, then we need a conceptual framework as well as a deductive system which will lead us to focus on these processes. Further we require a theoretical system in which both the economic decision behavior of individuals and groups can be explained. If such a theoretical system is to generate testable hypotheses of lasting significance it must allow for the variety of observable differences in individual behavior. Further, the theoretical schema should be such that the decision processes themselves can be expressed in a variety of content languages where each of these descriptions has the same classes of observable phenomena.

For example, if we are describing the decision processes of an individual we need a language that will accommodate such psychological characteristics as are involved in human learning; while if we are discussing

^{8/} The mathematical system is described in some detail in Chapter 3.

a larger economic system it may be more convenient to talk about different types of adaptive processes. As long as these processes are described in such a way that they can be tested against the same classes of data then hypotheses tested in one context can be related to those of another.

One possible approach to this general task would be to assume that each individual economic actor, whether individual or firm behaves according to its own, "rational" or "irrational" decision process. If each process is unique unto itself then the only way we can begin to understand the interactions among these processes is by positing the existence of some simple decision rules to account for some of the aggregate characteristics of the observed behavior. Such a procedure, however, will lead us right back to the classical position since the class of maximizing decision rules are precisely of this type.

Another, and if it is corroborated by empirical test, more powerful approach is to postulate that there is an invariance in the decision processes of various classes of economic actors. If a theoretical statement of these invariances is possible, then any one individual's decision processes are explainable by the combination of the invariant process and a specific set of parameters and processes that are particular to the individual, where these latter parameters and processes are directly related to observables or have operational forms of measurement. If it is possible to describe an invariant structure for individual decision processes, then the next step is to identify an invariant structure for the decision behavior of groups or organizations by developing a set of correspondences between the structure of individual and group processes. With such a theory, individual as well as group (firm, organization, and market) behavior would be explained on the basis of a set of invariant decision structures with the addition of a specific sets of observable

parameters and decision processes.

In order to construct such a theory the several components must be empirically specified. That is to say, the invariant structures must be identified and defined, and it must be demonstrated that these structures are sufficient to generate a wide variety of observed behavior. Moreover, to account for a particular stream of individual behavior techniques need to be developed which permit the specification of those parameters and processes which must be included in the statement of the individual's decision process. Once these structures and processes have been described they must then be submitted to empirical test. For this approach, like any other in science, can only be justified by its ability to lead to the development of theories which can withstand a process of refutation by empirical test. Consequently, it is toward a detailed examination of the empirical as well as the economic relevance of this conceptual framework that the remainder of this book is directed.

Chapter 12

FOUNDATIONS OF BEHAVIORAL THEORY

The discussion in the previous chapter argues that in order to construct testable theories of economic behavior it is first necessary to understand and to be able to explain the decision behavior of economic actors. While classicists and econometricians can undoubtedly agree with this statement, it has been demonstrated that their theories cannot be employed to establish the necessary explanations. One of their principal failings is that these theories cannot be submitted to test independently of specific economic contexts. Accordingly, if a new body of theory is to be constructed which can be corroborated by empirical test the implication is that it must be formulated in such a way so that at least some of its hypotheses are independent of a particular economic context. That is to say, if economics is to rest on a testable theory of economic processes, hypotheses about these processes must be corroborable by direct reference to a wide variety of observable behavior. In brief, a theory of economic processes is needed which is capable of satisfying several requirements. First of all, it must enable one to account for observed decision behavior occurring at a particular time and under specific conditions. Concurrently, it must allow one to be able to explain the decision behavior of individuals as well as groups of organizations. For if the behavior of firms or groups of consumers is to be explained, a theory is required which encompasses the decision behavior of such collections of individuals.

An understanding of economic decision processes appears to be the principal requirement to be satisfied if a science of economics is to be developed. Given this basic position, it follows that the search for

such a foundation should be conducted amongst the recent researches in behavioral and psychological theories of decision-making behavior.^{1/} Even though these theories are not all concerned with the same economic phenomena, they all employ a basic set of hypotheses which posit the existence of certain empirical regularities in the decision processes of economic actors. These hypothesized regularities are in turn derived from various researches in the simulation of individual decision behavior.^{2/} Accordingly, before one can accept behavioral theories as a possible basis for a science of economics it is necessary to be sure that the principal hypotheses of the underlying theory of individual behavior are both capable of test and have been corroborated by a number of tests.

^{1/} That this is not a novel suggestion is evinced by the fact that several economists, for example, A. Papandreou, "Some Basic Problems in the Theory of the Firm," in B. F. Haley, (ed.), Survey of Contemporary Economics, Irwin, Homewood, Vol. 2, 1952, pp. 183-219; and E. Grunberg, "Notes on the Verifiability of Economic Laws," Philosophy of Science, Vol. 24, 1957, pp. 337-348, have pointed out both the desirability and the possibility of reducing economic theory to psychological or behavioral terms.

The most notable of the early researches in behavioral theories as well as a detailed presentation of a behavioral theory of the firm are to be found in: R. M. Cyert and J. G. March, A Behavioral Theory of the Firm, Prentice-Hall, Englewood Cliffs, 1963.

^{2/} For a survey of research on simulation see the papers presented by G. H. Orcutt, M. Shubik, and G. P. E. Clarkson and H. A. Simon in "Simulation: A Symposium," American Economic Review, Vol. 50, Dec. 1960, pp. 894-932. For a more extensive analysis of the problems of simulating human decision-making behavior see the papers included in: E. A. Feigenbaum and J. Feldman, (eds), Computers and Thought, McGraw-Hill, 1963, Part 2, "Simulation of Cognitive Processes, pp. 269-380.

1. A Theory of Individual Decision Behavior^{3/}

The theory upon which the above noted theories of economic decision processes are based was developed to account for the problem solving behavior of individual subjects as they performed a number of specified tasks.^{4/} The purpose of the theory is to explain the process of human problem solving by identifying the classes of decision processes which are employed by humans while deriving the solutions to a range of different problems. Questions about problem solving behavior could, no doubt, be answered at several levels and in varying amounts of detail. This theory seeks to explain such behavior in terms of a set of basic information processes. These processes are partially defined by the theory's main postulates which state that for each problem solver there exists:

- "(1) A control system consisting of a number of memories which contain symbolized information and are interconnected by various ordering relations...
- (2) A number of primitive information processes which operate on the information in the memories....
- (3) A perfectly definite set of rules for combining these processes into whole programs of processing..."^{5/}

^{3/} This section is principally indebted to G. P. E. Clarkson and W. F. Pounds, "Theory and Method in the Exploration of Human Decision Behavior," Industrial Management Review, Vol. 5, Fall, 1963, pp. 17-27.

^{4/} The earliest statement of the theory is to be found in: A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a Theory of Human Problem Solving," Psychological Review, Vol. 65, 1958, pp. 151-166. A more recent statement, which includes some of the empirical investigation of the theory, is presented in: H. A. Simon and K. Kotovsky, "Human Acquisition of Concepts for Sequential Patterns," Psychological Review, Vol. 70, 1963, pp. 534-546.

^{5/} A. Newell, J. C. Shaw, and H. A. Simon, ibid., p. 151.

From these postulates it is clear the theory assumes that decision processes can be isolated as well as identified. Indeed, the theory also assumes that they can be represented by a series of straight-forward mechanical processes. In other words, the theory posits that decision processes consist of certain specific components, e.g. the memory, the basic information processes, and the rules for combining these processes into whole programs of information processing, which in turn are composed of collections of simple, describable mechanisms.

In order to clarify the empirical meaning of these postulates consider the following application of the theory of human problem solving to the decision processes of an investor of trust funds in a bank.^{6/} This theory of investment behavior was developed to account for the portfolio selection process of a particular trust investor. The basic postulates state that the trust investor has:

- (1) A memory which contains information associated with the general economy, industries, and individual companies. The information is ordered in associated lists. Although all investors may not associate a particular company with a given industry, the process of classification by industry is the primary basis for listing companies in the memory. The information related to each company may also vary among investors, but each company is represented as having a list of attributes with their values stored in the memory, e.g. growth rates of sales and earnings, price earnings ratio, dividend rate, etc.

^{6/} For a complete statement of the theory see: G. P. E. Clarkson, Portfolio Selection: A Simulation of Trust Investment, Prentice-Hall, Englewood Cliffs, 1962.

(2) Basic information processes which perform the tasks of searching the lists of information in the memory, selecting those items which have the required attributes, regrouping the selected pieces of information into new lists, and performing algebraic operations when necessary.

(3) A set of rules or criteria which determine the decision-making process by denoting the order and manner in which each information process is to be employed. This set of rules constitutes the structure of the investor's portfolio decision process.

As a further application of the basic postulates consider the theory of human problem solving which has been proposed under the name of General Problem Solver.^{7/} The object of this theory is to explain the problem solving behavior of individuals when they are involved in the solution of tasks for which means-ends analysis is an appropriate method of attack. In order to operate within the context of a particular problem situation the basic postulates of the theory require the following information to be provided:

For the memory:

"(1) A vocabulary, for talking about the task environment, containing terms like: object, operation, difference, feature...

^{7/} A. Newell, J. C. Shaw, and H. A. Simon, "Report on a General Problem Solving Program for a Computer," Proceedings of the International Conference on Information Processing, UNESCO, Paris, 1959, (Reprinted in Computers and Automation, Vol. 8, 1959).

(2) A vocabulary, dealing with the organization of the problem solving processes, containing terms like: goal type, method, evaluation....."

For the decision processes:

"(3) A set of programs defining the terms of the problem solving vocabulary by terms in the vocabulary for describing the task environment.

(4) A set of programs (correlative definitions) applying the terms of the task - environment vocabulary to a particular environment..."^{8/}

Within the context of a particular subject area, e.g. chess, symbolic logic, or trigonometry, GPS is a theory of human problem solving which essentially consists of a collection of general rules and detailed techniques for generating problem solutions. Because these processing rules are largely independent of the subject matter of a particular problem, e.g. capturing a bishop, proving a theorem, or proving an identity, GPS is more than a theory of one individual's decision processes. It is in fact the beginnings of a general theory which when suitably interpreted is sufficient to account for the decision behavior of a number of individuals.^{9/}

^{8/} A. Newell, J. C. Shaw, and H. A. Simon, Op. cit., (Computers and Automation), pp. 11-12.

^{9/} A discussion of the conditions under which these theories are subjected to test and an examination of the evidence currently available is left until later.

As can be seen from these two examples the theory of human problem solving asserts that the decision processes of individuals can be analyzed and described in terms of information processes. When these operations are collected into a set of statements which describe the behavior of the individual or individuals under investigation, such statements become a theory of the decision-making process. That such a set of rules can be considered to be a theory is evinced by the requirement that it must be possible to deduce unequivocally the externally observable behavior which will be generated by it. To ensure the satisfaction of this condition, the theory is translated into a formal language (in this case a suitable computer language, about which more will be said below) and the logical consequences are derived by performing each operation according to the specified rules.

2. Goals and the Structure of Decision Processes

From this discussion of the basic postulates and assumptions of the theory of human problem solving it is now possible to examine the manner in which observed behavior is to be classified and structured. According to the theory all decision behavior can be analyzed and described by a set of processing rules operating on a specific collection of information which is available to the decision-maker. This information is available to the individual either in his memory or in his environment. But before the theory can be usefully applied to a particular situation it is necessary to be able to isolate and identify the principal decision processes as well as the structure by which they are related.

Most theories of human behavior include a reference to the purpose or goal toward which, it is argued, the behavior is directed. In classical economic theory the goal of the consumer is to maximize his utility (expected) subject to his budget constraint. Similarly, the goal of the

firm is to maximize net revenue, or in the case of a recent proposal the goal is to maximize net sales subject to a profit constraint.^{10/} While disputes may arise over which goal the behavior is supposed to serve^{11/}, most theories reflect the general belief that behavior can be usefully described in these terms.

Under the theory of human problem solving a specific stream of observed behavior is described and explained by identifying a particular set of decision rules as well as the information upon which they operate. Within the context of this theory an external goal or purpose is not relevant to the understanding of the behavior. For the behavior of a set of mechanisms operating in a particular environment determines the consequences of final output. Behavior is generated by specific processes operating on items obtained from the memory or the environment and is not a function of external goals.

To help clarify the point consider the following examples of "goal directed" behavior. To begin with consider an inanimate torpedo. Suppose for the moment that it has been constructed in such a way that its steering mechanism is directly connected to an electronic mechanism which is sensitive to certain vibrations in the water. Under normal conditions this electronic mechanism will process the incoming vibrations and alter

^{10/} W. J. Baumol, Business Behavior, Value and Growth, Macmillan, New York, 1959.

^{11/} For example, consider the list of different goals to which individuals are posited as striving towards in gaming and bargaining situations, e.g., R. D. Luce and H. Raiffa, Games and Decision, Wiley, New York, 1957; R. D. Luce, Individual Choice Behavior, Wiley, New York, 1959; and T. C. Schelling, The Strategy of Conflict, Harvard University Press, Cambridge, 1960.

the direction of the torpedo in conformity with these signals. If we, as observers, witnessed this torpedo intercept a moving object on the water, we might describe it as a homing torpedo but we would be most unlikely to ascribe to it the goal or purpose of destroying particular types of floating objects. The behavior of the torpedo at any instant is completely described by a knowledge of its control process and the incoming signals. While the torpedo may or may not eventually strike a floating object, the inclusion of this result of its behavior is not relevant to the explanation of its behavior when it is still some distance from the object.

As a second, and animate example, consider the problem of describing the behavior of an investor who is selecting a portfolio for a client. One of the first items to be determined is the investment policy for this account. Once the policy is selected, it can be applied to a suitable list of securities in order to determine which securities are to be included in the portfolio. If the policy is "growth" then a decision process is needed which will select a particular set of growth stocks from the total list of such stocks which are available at the time. We, again as observers of this process, might describe this selection process as one which seeks to select a growth portfolio or one which has growth as its goal. But, in fact, the actual growth rate of the resulting portfolio is largely independent of the process by which it is selected. As a result, the term "growth" is really the name for the process which acts as the selection mechanism. This is not to say that the investor could not have a "target" rate of growth, say ten per cent per year. If at the end of a year the portfolio has not grown in value by this amount such a failure may well trigger off a re-examination of the existing portfolio. However, even though this target rate of growth may be viewed by an outsider as a goal to which the investor is striving, it is on closer

inspection no more than a control device--one which under certain conditions activates certain other processes, such as searching for other securities.^{12/}

A third example of the manner in which the theory of problem solving employs the term goal is given by GPS itself. GPS, as mentioned above, is a theory of problem solving which encompasses problems to which means-ends analysis is appropriate. Hence, GPS is able to work on problems which can be formulated in terms of objects and operators. An operator is a decision process or a process for developing a decision process which can be applied to certain objects to produce different objects. An object is described by its features and one of the commonest features that distinguish pairs of objects is the differences between them. In order to address itself to a specific problem GPS employs three types of goals: A Transform goal, a Reduce Difference goal, and an Apply Operator goal. To each of these goals is associated one or more methods for achieving it. Consequently, when a goal is activated the relevant methods for accomplishing this goal are brought out from the memory and tried. For example, one method of changing the object a into the object b is to note the difference between them d and by activating the Reduce Difference goal try to find a method which when applied will reduce differences of this sort. If a method is found it is applied by the Apply Operator goal.

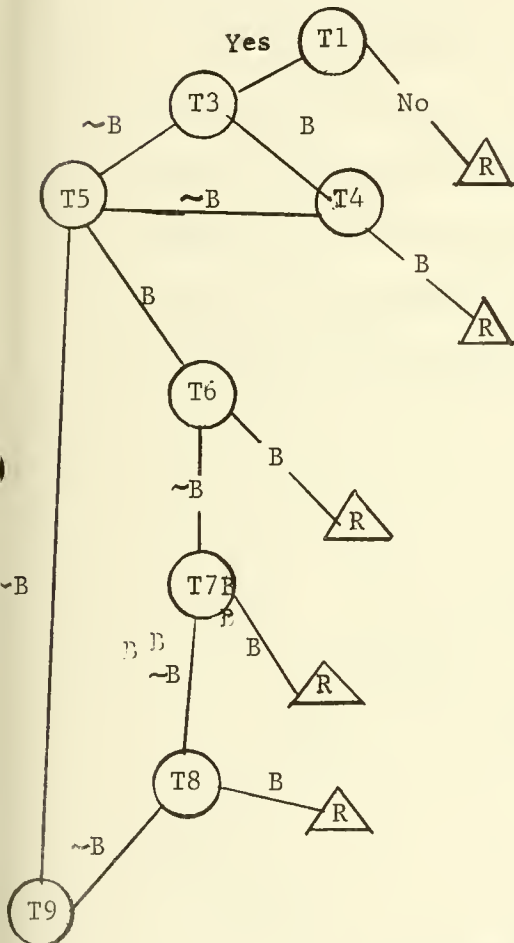
^{12/} For an excellent analysis of goals and their effect upon decision behavior see: W. F. Pounds, "A Study of Problem Solving Control," unpublished Ph. D. dissertation, Carnegie Institute of Technology, 1964; available in mimeograph, School of Management, Working Paper 33-63, Massachusetts Institute of Technology.

Once again, it is clear that the term goal refers to the name of a decision process and not to some result or consequent which is external to the decision process. Accordingly, this approach to decision behavior argues that to understand decision behavior only requires one to discover the decision processes which determine the observed behavior. This is not to say that humans do not "have purposes" which affect their behavior, such as wanting to get married, desiring great wealth, prestige, or a happy life. On the contrary, these objectives or targets frequently occur in conversations and no doubt influence the motivations and emotions of many individuals. However, in order to describe and explain an observed stream of problem solving behavior it is not necessary to know the source of inspiration, frustration, or motivation; it is sufficient to be able to describe the processes and the concomitant information which determine the observed behavior. To command someone "to do better" is the height of futility unless that person has or is able to acquire a set of decision processes which will lead him to produce the desired results. Behavior is determined by decision processes and to stipulate an objective without providing the requisite decision rules and control processes is not the way to produce the required behavior.

Decision processes which select or operate on the information in the memory or environment are represented, under this theory of decision behavior, by nets. A net is an associated list of tests or filters through which the information passes. Each test or item in the net is the name of another process, and the behavior of the entire process is the result of the net operating upon the information that passes through it.

For example, in the theory of investment behavior the decision processes or discrimination net which selects the individual securities for a specific portfolio is represented by a collection of tests through

which a security must pass if it is to be accepted. Each of these tests may be simple or complex, but the discrimination net itself will only contain their names and the order in which they are associated to one another. In the following net, which is part of the Growth Portfolio discrimination net, T1-T9 represent a particular sequence of tests that are applied in turn to an appropriate list of securities.^{13/}



Dictionary

- T1 - Mean growth in price (past) $\geq 20\%$
- T3 - Mean growth in earnings per share (past)
- T4 - Mean growth in sales past
- T5 - Forecasted growth in earnings per share (1 year)
- T6 - Forecasted growth in sales (1 year)
- T7 - Mean growth in cash flow per share (past)
- T8 - Mean growth in profit margin (past)
- T9 - (y) on Relative Value List
- B - "Below" average for industry
- ~B - "Not Below" average for industry
- R -- Reject.

FIGURE 1

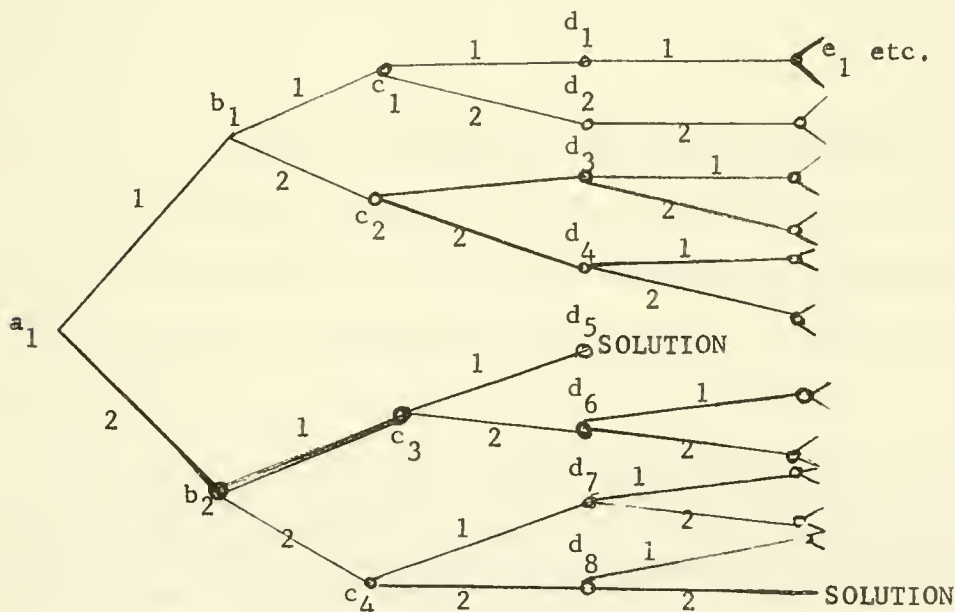
In this net processing begins with the test named T1. If a security passes this test, T6 is applied. From T3 the security will either be processed

^{13/} For complete description of this net and the way in which it is employed see: G. P. E. Clarkson, Portfolio Selection, Op. Cit., Chapter 4.

by T4 or T5 depending on the outcome at T3. Aside from T1, the outcome of each test depends on the relative characteristics of each security. Further, the tests are arranged in hierarchies so that if a specific security is "below average" with respect to the characteristic examined by T5 it must pass through T6, T7, and T8 if it is not to be rejected and is to return to T9 and the remainder of the net.

A further example of this structure of decision processes is provided by employing a maze as a representation of the problem solving process. A maze is an hierarchical structure of paths, (See Figure 2) some elements of which belong to the set of "correct paths"--i.e., they lead to the solution of the problem. The maze can be represented as consisting of all the possible paths which could have been taken.^{14/} Or the maze can

FIGURE 2



^{14/} In this case the maze is analogous to the notion of a game tree used in game theory and statistical decision theory, see: R. D. Luce and H. Raiffa, op. cit.; and H. Raiffa and R. Schlaifer, Applied Statistical Decision Theory, Wiley, New York, 1954.

represent that set of paths taken by a single individual to reach a particular solution, as in the investment example above.^{15/} Under this interpretation all problem solving behavior can be represented by a sequential list of operations. Since discrimination nets have the required associative and hierarchical structure, all decision processes can be represented by discrimination nets. Consequently, in order to be able to empirically identify a specific decision process it is necessary to know the contents of the tests or processes as well as the way in which they are interconnected in the net. Once these items are known the behavior of the decision process is fully determined. For by hypothesis, the generated behavior is the result of the decision process acting on the information stored in the memory or the environment. As a result, the key to the explanation of observed decision behavior lies in the ability to isolate and identify the contents of discrimination nets, and, as a consequent, the information required by these nets.

3. On the Explanation of Decision Behavior^{16/}

In the preceding sections it has been suggested that the theory of problem solving behavior is sufficient to provide the empirical foundations for the explanation of observed decision behavior. Since all theories claim

^{15/} For a more detailed discussion of the maze as a prerepresentation of the problem solving process see: A. Newell, J. C. Shaw, and H. A. Simon, "The Processes of Creative Thinking." in H. E. Gruber, G. Terrell, and M. Wertheimer, (eds.), Contemporary Approaches to Creative Thinking, Atherton Press, New York, 1962, pp. 63-119.

^{16/} This section is in part drawn from, G. P. E. Clarkson, "Verification and the Function of Laws in Micro-Economics," Industrial Management Review, Vol. 4, Fall, 1962, pp. 41-58 and from The Theory of Consumer Demand, op. cit., Chapter 7.

to be able to explain something and, as has already been shown, not all theories do so, it is appropriate to re-examine briefly what is meant by the word "explain."

To establish a scientific explanation ~~for~~ the occurrence of an event, three conditions must be satisfied.^{17/} The first is that the occurrence of the event must be deducible as a direct consequence from the conjunction of the theory and the appropriate initial conditions. For this condition to be satisfied the theoretical system must conform to the general rules of logic which govern the formation and manipulation of deductive systems. Theories which are stated in verbal or mathematical form can meet these conditions just as well as theories stated in terms of a computer program. In ~~all~~ cases the theory can be constructed so that the process of deductive inference will conform to the general rules governing deductive systems.^{18/}

The second condition is that the theory itself must contain at least one general hypothesis or law which has been confronted with and survived a process of refutation by empirical test. Accordingly, at least one of the theory's hypotheses must be stated in such a manner that it can be corroborated by empirical test. The third condition requires the statements describing the initial conditions to be empirically true.

If the theory of human problem solving is to provide the empirical foundations for testable theories of decision behavior, then the explanations provided by such theories must satisfy these three conditions. From the previous sections it is clear that an explanation of observed behavior is achieved by applying the hypothesized decision processes to the information

^{17/} For a further discussion of scientific explanations see Chapter 2, and the references cited there.

^{18/} These rules are described and detailed references are provided in Chapter 2.

(initial conditions) contained in the memory or the environment. If the generated behavior matches the observed (in a manner to be discussed in the next chapter) then that set of observed behavior is said to have been explained.

In order to determine whether such explanations satisfy the three criteria consider the following example of an explanation of an economic event that is proposed by the theory of trust investment mentioned above. The event to be explained is the selection of a portfolio of securities by a particular trust investor for a specific trust account. To establish the explanation of this event the theory requires the initial and boundary conditions to include: Data on the historical, current and expected values of relevant financial attributes, e.g. price, yield, earnings per share, profit margin, growth rates of price, sales, and earnings, etc. for a specified list of securities; data on the historical, current and expected values of specific industrial and economic indicators; and certain data on the particular trust account in question. The hypotheses of the theory are concerned with the trust investor's decision process. They posit that the decision process can be represented by: (i) a memory which contains the data noted above listed in a particular form; (ii) a set of procedures which allow the data in the memory to be searched, manipulated, and desired items selected for further processing; and (iii) a set of decision rules that determine the decision-making process by stipulating when and where each decision process is to be carried out. These hypotheses include statements about the structure of the memory and of the individual decision processes, the way in which expectations are formulated, and the sequence in which the various decision processes are applied.^{19/} In

^{19/} It should not be forgotten that these hypotheses are stated in sufficient detail to permit their programming and testing on a digital computer.

brief, the hypotheses define both the structure and the order of the decision processes. When they are employed in conjunction with the initial and boundary conditions the decision processes select a specific portfolio. Accordingly, the explanation of the selection of a particular portfolio is established by applying the decision procedures given by the theory to the data of the security market and the specific trust account in question.

Since the statements describing the initial and boundary conditions refer exclusively to observables the third criterion is manifestly satisfied. Moreover, the theory itself is translatable into an unambiguous computer program. Thus, as long as the computer language contains all the requisite properties of a formal language, the deductive system (the computer program) satisfies the first condition. Hence, in order to determine whether this explanation can be considered to be "scientific" all that is necessary is to show that the theory contains at least one general hypothesis which is both refutable and not yet confuted by empirical test.

To demonstrate that this criterion is satisfied all one need do is submit the theory of trust investment to a number of empirical tests. If the three postulates of the theory of human problem solving are taken as exemplars of the general hypotheses for the theory of trust investment, then a successful series of empirical tests on these hypotheses will constitute evidence for their empirical validity. Indeed, since the initial conditions are all observable, all such tests constitute potential disconfirmation.

For example, the theory of trust investment was constructed by incorporating into its hypotheses such decision processes as were observed (and inferred) from the trust investor's behavior. To test the theory's ability to reproduce the decision behavior the initial conditions were specified by providing the requisite data of the security markets and of

some specific trust accounts for a particular time period. The theory was then required to generate portfolios for these trust accounts. The specific portfolios, however, only constitute the end product of the investor's decision process. Consequently, the theory was also subjected to a set of tests which compared the behavior generated by its hypotheses with the recorded decision behavior of the trust investor.

Although it is of some interest to be able to explain and predict the actual portfolio selections of a specific trust investor, there are presumably a variety of theories which will accomplish this result. What is of much greater importance, is the fact that in a number of actual tests the theory's decision behavior compared favorably with that of the trust investor--i.e. it appeared even on close inspection, that the theory was employing similar decision procedures and was arriving at the same results for substantially the same reasons as the particular investor under investigation. By comparing the behavior generated by each of the theory's major hypotheses directly with the observed it is possible to submit the theory's decision processes to a process of refutation by empirical test. Manifestly, this testing procedure can be repeated. Also the hypothesized processes can be compared with observed behavior to whatever level of detail is appropriate or desired. As a result, while the theory of trust investment may or may not have been adequately confirmed, it is demonstrably possible to corroborate its principal hypotheses.^{20/}

^{20/} The actual procedures by which such hypotheses can be submitted to test are discussed in the next chapter. For a detailed presentation of the results of the direct tests of the investment theory's hypotheses see: G. P. E. Clarkson, Portfolio Selection, op. cit., Chapter 7.

4. Towards a General Theory of Decision Behavior

Once it is possible to corroborate hypotheses about decision processes within a given economic context the next important step is to determine whether some of them can also be subjected to test independently from their application in, say, the theory of portfolio selection. For, although individual economic decision processes are of interest by themselves, one cannot establish the empirical validity of a general theory of decision processes unless it is possible to subject some of the hypotheses to test independently of a specific economic context.

Behavioral theories, just as theories of individual decision behavior, are concerned with explaining various aspects of human decision-making behavior. All of these theories contain hypotheses which make definite assertions about the structure and ordering of the relevant decision processes. Since each of these theories deals with various aspects of observed decision behavior it is manifestly both possible and practicable to study human decision-making behavior in a diver's (non-economic) number of contexts. For example, the decision behavior of individuals engaged in the solution of problems in geometry, logic or chess could be used as the framework within which to test the empirical validity of many of the hypothesized decision processes.^{21/} It is not being suggested that all

^{21/} Indeed, the basic hypotheses employed in the theory of trust investment were derived from researches in individual decision behavior which used the problems of geometry, logic and chess as their sources of empirical evidence. See for example: A. Newell, J. C. Shaw, and H. A. Simon, "Empirical Explorations of the Logic Theory Machine," Proceedings of the Western Joint Computer Conference, February, 1957, pp. 218-230, and "Chess-Playing Programs and the Problem of Complexity," IBM Journal of Research and Development, October, 1958, pp. 320-335; and H. L. Gelenter, J. R. Hansen, and D. W. Loveland, "Empirical Explorations of the Geometry Theorem Machine," Proceedings of the Western Joint Computer Conference, 1960, pp. 143-159. (All three articles are reprinted in E. A. Feigenbaum and J. Feldman (eds.) Computers and Automation op. cit., pp. 109-133, 39-70, 153-163 respectively.)

hypotheses of a single theory, say the theory of trust investment, can be tested independently from its economic context. Certainly, some of them, e.g. the criteria by which companies are listed in the memory, the order in which the testing and processing of the individual securities, etc., are performed will be peculiar to the specific economic context. What is being asserted is that there are a certain number of invariances among the decision processes of different problem solvers, and that it is possible to test for their empirical truth value in a variety of empirical contexts.

Consider, in this respect, the theory of human problem solving. It contains three postulates which assert the existence in a human decision-maker of a memory, some primitive information processes, and an hierarchy of decision rules. The theory of trust investment, like the General Problem Solver, turns these postulates into testable hypotheses by specifying in detail the content and structure of the memory, the information processes, and the content and order of the decision rules. If it were not possible to specify how to characterize and empirically interpret these processes, then it would not be possible to directly adapt these postulates into a testable theory of individual behavior. Moreover, unless invariances, like the structure of the contents in memory and the structure of the decision processes themselves, exist among decision-makers it is not possible to construct theories of decision behavior in this manner.

Implicit in this last statement is the postulate that invariances exist among the decision processes of different problem solvers. Indeed, it is being posited that these invariances not only exist but they can also be isolated, identified and empirically confirmed. As evidence for this postulate consider the number of theories of human decision behavior which

are directly derived from the theory of human problem solving.^{22/} While it is not being suggested that this postulate can be accepted as a well tested empirical regularity, it is clear that it has been subjected to a number of empirical tests. Further, and what is perhaps more important, it is in principle possible to submit such hypotheses to test and the references to the literature point to examples where such a program is already being carried out.

In order to demonstrate the empirical possibility of developing testable theories of decision behavior, it is sufficient to show that it is possible to implement such a testing procedure. For, if the appropriate tests are performed and some of the hypotheses are not disconfirmed, then these relations will have become the independently tested empirical regularities which will constitute the empirical foundations. Once a set of empirical regularities is established then all hypotheses that can be deduced from them either alone or in conjunction with other postulates become in turn capable of being, at least indirectly, confirmed or disconfirmed by empirical test. Consequently, when a particular theory, say the theory of trust investment, is employed in the explanation of a specific economic event, the independent testability of its principal hypotheses ensures that a scientific explanation has been established.

Concurrently, once some hypotheses about decision behavior have been established as empirical regularities the prediction of the occurrence of an

^{22/} For examples of these theories see: E. A. Feigenbaum, "The Simulation of Verbal Learning Behavior," Proceedings of the Western Joint Computer Conference, May, 1961, pp. 121-132; J. Feldman, "Simulation of Behavior in the Binary Choice Experiment," Proceedings of the Western Joint Computer Conference, May, 1961, pp. 133-144; (Both of these papers are reprinted in E. A. Feigenbaum and J. Feldman (eds.) op. cit., pp. 297-309, 329-346 respectively); R. K. Lindsay, "Inferential Memory as the Basis of Machines which Understand Natural Language," in E. A. Feigenbaum and J. Feldman, op. cit., pp. 217-233; and A. Newell and H. A. Simon, "GPS, A Program that Simulates Human Thought," in E. A. Feigenbaum and J. Feldman, op. cit., pp. 279-293.

event can also be employed as the basis from which to test part or all of the hypotheses in a theory. For example, the trust investment theory was subjected to test by requiring it to predict, under different market conditions, the investor's portfolio selections. In this case, the theory's portfolios can be used to determine whether the theory's decision processes are sufficient to reflect the changing economic conditions in the securities which are selected. If the predicted portfolios compare favorably with the investor's under one set of conditions but not under another, then it would follow that the decision processes were not sufficient to permit the theory to adapt its selections to the prevailing market conditions.^{23/} Therefore, to the extent that empirical regularities of decision-making behavior can be established it is then possible to develop general theories of decision behavior which can explain and predict the observed behavior of a variety of economic actors.

^{23/} The evidence on this point is presented in ibid., Chapter 6.

Chapter 13

Some Problems of Application

Imagine, for the moment, that we, you and I dear reader, wish to develop a theory to explain a particular sequence of observed behavior. The behavior in question is of such frequent occurrence that a theory which is sufficient to explain it will considerably improve our understanding of the behavior of the economic factors involved. Moreover, for purposes of social welfare we should like to be able to exercise some degree of control over this form of economic behavior. As a result, an understanding of the decision processes involved is clearly of great importance to us. For, once we have a theory which can explain the observed behavior, we will also have an empirical basis from which to discuss and experiment with alternative methods of control.

From the previous two chapters we are convinced that the theory must account for the economic decision processes of the individuals involved. Since no other alternatives seem to be available it appears that we should construct our theory upon the foundations provided by the theory of human decision behavior. That is to say, if our appreciation of the situation is correct what we need to do is take the general theory of decision behavior as our theoretical base and by adding the appropriate information and decision rules develop a testable theory of the behavior in question. Because all decision processes can be represented by ordered structures of information processes our task is quite straight forward--it is to isolate and identify both the requisite information and decision processes. Once our hypotheses are formulated and the theory is constructed the next step is to subject the theory to empirical test. If the theory survives the test or tests our task is complete. For with this theory we can now explain the behavior which stimulated our interest and this theoretical activity. Manifestly, the principal

components of such an endeavor consist of first developing the individual hypotheses and then submitting these hypotheses to a process of refutation by empirical test. This chapter is directed toward an examination of both of these processes.

1. On the Construction of Decision Theories^{1/}

The theory of human problem solving posits the existence of a memory, a set of information processes, and a program of processing rules. Thus, if theories of decision behavior are to be based upon this foundation then behavioral theories must include these three postulates as part of the total set of hypotheses. Further, in order to construct a theory to explain a particular set of behavior it is necessary to specify the empirical interpretation of these postulates in complete detail. It is not sufficient, for example, to postulate that the economic factor in question has a memory. For just any memory will not serve the purpose at hand. Before the postulate about the structure of memory has empirical meaning the interpretive rules must specify both its content and the order in which the items are associated to one another.

In the theory of trust investment the postulate about the structure of the memory is given empirical meaning by a number of interpretive rules. First, since the theory deals with that part of the investor's memory relevant to the portfolio selection process, the primary criterion by which companies are listed in memory is noted. Once the companies are specified as being ordered by industry the appropriate attributes of each company as well as their values need to be discovered. These attitudes and values are associated by a particular memory

^{1/}This section is in part indebted to G. P. E. Clarkson and W. F. Pounds, op. cit

structure to each of the companies that are being considered. Since the theory also includes information on industry as well as general economy indicators, these items must also be related by a specific structure to the information already placed in memory. In the trust investment case the detailed information is drawn from observations on one individual. However, the general requirement for the specificity of the contents of memory must be satisfied whether the theory concerns the behavior of one or many individuals. For the contents of the memory and the relevant items in the environment (which can frequently be represented as a part of the memory) constitute the initial conditions. Unless, the initial conditions of a theory are both specifiable and empirically observable it is not possible to submit the theory to empirical test. Consequently, the identification of the structure and contents of memory is the first important step in the construction of a theory of decision behavior.^{2/}

The second main hypothesis asserts the existence of a set of primitive information processes which operate on the information already located in the memory. While there are undoubtedly a number of ways in which these information processes could be specified the representation employed by this theory is defined by the language that is used--^{3/}. IPL V is a formal language which satisfies the syntactical rules governing languages in deductive systems. Accordingly, a theory stated in this language is able to satisfy the formal requirements of a

^{2/}For detailed investigations of memory structures see: E.A. Feigenbaum, op. cit. and R. K. Lindsay, op. cit.

^{3/}This language is presented in complete detail in A. Newell, (ed) Information Processing Language V Manual, Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

deductive system^{4/} and can constitute the language for a scientific theory. The language itself is composed of a set of basic information processes and a number of interpretive rules for executing the information processes. The information processes, denoted by the prefix J, are in turn based upon an hypothesized structure of the memory. As a result, these processes are principally concerned with finding, deleting, adding, re-ordering, and manipulating the items associated in list structures of the memory. These specific processes are the empirical interpretations of the general class of primitive information processes. Even though it is not asserted that they represent a complete set of such processes, to employ this language is to adopt the hypothesis that these processes are sufficient for the purpose at hand. That is to say, in order to empirically specify a theory's hypothesized decision processes this language provides a sufficient set of primitive information processes so that either by themselves or in appropriate combinations they are the requisite interpretive rules. Without such a set of interpretive rules hypotheses about decision processes would be devoid of empirical content. Despite the fact that these primitive processes could be specified in a number of different languages, the actual existence of IPL V provides the necessary assurance that hypothesized decision processes can be empirically interpreted.^{5/}

The third postulate claims that observable decision behavior is a consequence of a set of rules or decision processes which combine the primitive

^{4/}For further discussion of these requirements see Chapter 2, sec. 1 and the references cited there particularly: G.P.E. Clarkson, The Theory of Consumer Demand, op. cit. Ch. 2.

^{5/}A number of these points are discussed with reference to particular examples of human decision processes in: H. A. Simon and K. Kotovsky, op. cit.

information processes into whole programs of processing. As has already been noted, a theory of decision behavior is a statement of the ordered structure of decision rules which describe the decision behavior under investigation. Consequently, in order to construct a theory of a particular stream of decision behavior, it is necessary to isolate and identify the decision rules which guide and constitute the decision-making process.

For example, consider once again the portfolio selection process of a trust investor. Since it is hypothesized that all decision behavior can be analyzed in terms of a set of decision routines which act upon a set of information contained in the memory, (the memory being the general repository of all pertinent information including that supplied by the environment) this theory represents the investment process as consisting of three major segments: (a) processes concerned with the analysis and selection of, from an initial set of stocks, a list of securities which are currently suitable for purchasing; (b) processes which determine the investment policy appropriate for each account; (c) processes which perform the actual selection of the individual securities for the portfolio for each account.

In accordance with the first postulate the information in the memory consists of ordered lists of data on specific economy and industry variables as well as data for a ten-year period on the relevant attributes of the total set of companies (eighty in this case) and their securities. Section (a) of the theory contains decision processes which employ these data to create various ratios and indices by which it will be possible for other processes to judge the relative performance and relative value of one company's stock against another. Data on

expectations are also included and are reduced by additional processes so that patterns can be found and recognized. A pattern recognizing process is then employed to create a list of stocks suitable for current acquisition. This list is derived from the original set of securities and its contents depend directly on the outputs of the relative performance, relative value, and expectational processes.

Section (b) of the theory consists of a set of decision processes which formulate an investment policy for each account. The investment policy is derived by processing certain data taken from the bank's records and the legal instrument setting up the trust account as well as data on specific attributes of the client or the trust account. The principal hypothesis of this decision process is a discrimination net which associates certain patterns of attributes and their values with specific investment policies.

In section (c) the portfolios are chosen by applying the selection processes ^{6/} associated with each investment policy to the list of securities generated by section (a). Concurrently, decision procedures are employed which determine how many shares to purchase of each security that is selected as well as how to ensure that the portfolio is appropriately diversified. The end result is a portfolio of securities for a specific trust account where the theory specifies the name of each security, the number of shares to purchase, the price per share at that time, and the total amount to be expended for each security.

From this brief description it is apparent that this is a moderately large and complex theory of decision behavior. In order to develop the three hypotheses

^{6/} These processes are represented by discrimination nets, an example of which is given in Chapter 12, p. 253.

of the theory of human decision-making behavior were empirically interpreted by adding the appropriate information and decision rules. Each decision process was constructed by observing and reviewing in detail a particular investor's decision behavior. Consequently, to construct such a theory it is important to know how to discover, collect and fit into the general structure the requisite information and decision rules. This may sound like a formidable task, but the general theory provides the structure with which the data are to be sorted and arranged as well as an outline which guides the observational process. Accordingly, even though the identification of the components of specific decision process requires careful observation, the task is made quite practicable by knowing what to look for.

Within this framework the task of constructing a theory to account for a particular sequence of observed behavior becomes a problem of uncovering the principal decision rules employed by the decision-maker. To obtain these data a variety of interview and observational techniques can be employed. The following list is merely an outline of some methods that have been used to advantage:

a) Interview

One method of discovering the components of an individual's decision process is by the question and answer approach of a normal interview. If the decision process in question is one which is frequently employed by the individual, questions about the procedure followed, the records consulted, the information that is processed, and the output can provide a rough outline of the more important parts of the decision process. Interviews are frequently more rewarding if there is one person to ask the questions while another takes notes. But, it must not be

overlooked that this approach in effect asks an individual to describe and in part justify why he behaves as he does. To the extent that many people are unable to describe in detail by what process they reached a particular decision the information gathered in this manner must be regarded with some caution.

b) Protocols of Decision Behavior

A more reliable guide to the identification of decision processes is by taking protocols of an individual's decision behavior. A protocol is a tape recorded transcript of the verbalized thoughts and actions of a subject who has been instructed to think or problem solve aloud. Consequently, the transcript is a record of the subject's thought processes while he is engaged in making a decision. Since a protocol is a detailed description of what a person does while problem solving it avoids some of the difficulties inherent in the interview and questionnaire techniques.^{7/}

c) Constrained Problem Solving Interviews^{8/}

A variant on the interview approach is to ask the subject to write out a decision process which he is willing to defend as being able to accomplish the task at hand. By requesting him to write out the decision processes and then

^{7/}"Thinking aloud is just as truly behavior as is circling the correct answer on a paper-and-pencil test. What we infer from it about other processes going on inside the subject (or the machine) is, of course, another question. In the case of the machine, the problem is simpler than in the case of the human, for we can determine exactly the correspondence between the internal processes and what the machine prints out," A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a Theory of Human Problem Solving," op. cit., p. 156.

The relevance of protocol data for testing purposes is discussed later in this chapter.

^{8/}For examples of the application of this technique see: W. F. Pounds, op. cit.

posing such questions as "but what happens if.....", he may be led to expand and alter what he had previously written down. Such modifications provide useful information on what are the important items in the decision process. Additional data can be obtained if it is possible to get the subject to employ his written decision routine to make one or more actual decisions. If, after observing the behavior of his own routine he is satisfied with its behavior, then this is a good basis from which to develop specific hypotheses concerning his decision behavior.

Throughout the data gathering process checks must be made to ensure that the relevant parts of the decision process are being identified. One way of checking initial hypotheses is to construct simple nets and decision rules. By applying these rules to the appropriate data one can readily determine whether they are going to be sufficient to reproduce the observations recorded, for example, in the protocols. If a record of past decisions is available hypotheses can be tested against these data as well. The object of this testing is, to identify the principal decision processes and data inputs which must be included if the observed behavior is to be explained. The construction of such a theory is, however, only the first part of the total process. Once a theory is built it must be tested. And the remainder of this chapter is devoted to an examination of this stage of the experimental procedure.

2. On Testing Decision Theories

One theory is a "model" of another theory only if their postulates and hypotheses are structurally similar.^{9/} Hence, a particular application of a

^{9/}For a stimulating as well as intensive examination of the relation between models and theories see: M. Brodbeck, "Models, Meaning and Theories," in L. Gross, (ed) Symposium on Sociological Theory, Row, Peterson and Co., 1958, pp. 373-403.

theory to a specific set of decision processes is a model of those processes. For example, when the general theory of problem solving behavior is employed to develop a theory of investment behavior, this application is a model of the general theory. Similarly, when a theory of trust investment is applied to a particular individual, the theory of the specific investor is a model of the investment theory. Moreover, as it is usually difficult to find general data against which to test general theories, theories are customarily submitted to empirical test by testing specific models against particular collections of data. In short, the process of testing a theory is in actuality a process of submitting a particular model of this theory to specific tests.

In order to examine the testing procedure, assume for the moment that one has at hand a theory of a particular set of decision behavior. Manifestly, the testing procedure for such a theory must take into account the fact that it is necessary to be able to check the final output as well as the decision processes by which the output was produced. Accordingly, the first step is to construct a specific model of the theory by specifying, where necessary, the particular parameter values (initial conditions) and decision rules that pertain to the context in which the theory is to be tested. Next, the model, i.e. the statements and decision rules which describe the behavior under investigation, and the statements containing the appropriate initial conditions are translated into a suitable computer language--e.g. Information Processing Language V. The computer is then activated and, as in the more familiar case of scientific theories, the logical consequences are derived by performing the individual operations according to the specified rules. Finally, to conduct

the test the behavior generated by the model is compared to the observed behavior of the individual or individuals under investigation. When the model yields results that are consistent with the observed, the theory is said to be sufficient to account for the recorded decision behavior.

Given such a model it is now possible to examine a number of problems that are raised by this testing procedure. To begin with what criteria are to be employed to discriminate between models that successfully reproduce observed behavior and those that do not? One answer is to accept the model as being corroborated when the results generated by it are consistent with those obtained from human decision-makers. In other words, accept the model, and hence the theory, when it is sufficient to account for observed decision behavior. Such an answer, however, does not provide an operational criterion for distinguishing when the results are to be considered "consistent." Unfortunately, there is no one criterion which can directly perform this service. As in any branch of empirical science it is not possible to "prove" that a theory or model is "empirically true." The best that can ever be said for a theory is that it has not yet been disconfirmed by empirical test. Accordingly, one cannot prove that a model of certain decision processes is empirically true. The best that can be done is to submit these models to more and more stringent tests in order to eliminate those hypotheses, and consequently theories, that are demonstrably false.

One testing procedure that meets this latter requirement is the adaption

of Turing's Test^{10/} proposed by Newell and Simon.^{11/} Turing was concerned with creating a test which would determine whether a machine could think. He called his test an imitation game and it proceeds as follows:

The game is played by three contestants--a machine, a human and an interrogator--and there are two channels of communication (say teletypes) which link the interrogator, separately to the human and the machine. The object of the game for the interrogator is to specify the identity of the two players. Active questioning by the interrogator is allowed, and the machine's task is to delude the interrogator while the human is supposed to want to do his best to reveal his "true" identity. The interrogator succeeds and the machine is declared unable to "think," if on a given number of trials he correctly identifies the players on a better than chance basis.

The adaption of Turing's Test to the problem of discriminating between the output of a specific model and the decision behavior of the human proceeds in a similar way: Data are gathered on the decision behavior of one or more subjects by making protocols or other records of the decision processes. The output generated by the model is also collected and can now be directly compared with the recorded human behavior. This comparison can be carried out at many levels of detail. The only restriction is the level of detail provided by the data on the human's decision processes. When the model produces behavior that meets the criterion of Turing's Test the model is sufficient to account for the

^{10/} A.M. Turing, "Can a Machine Think?" Mind, Vol. 59, October, 1950, pp. 433-460. (Reprinted in E.A. Feigenbaum and J. Feldman, op. cit., pp. 11-35).

^{11/} A. Newell and H.A. Simon, "The Simulation of Human Thought," in W. Dennis, (ed), Current Trends in Psychological Theory, University of Pittsburgh Press, 1961, pp. 152-179.

decision-making behavior under investigation. This test can be applied to the output of the model as a whole as well as to the behavior of the individual decision processes. In the former case the test might be considered to be quite weak since there are presumably a variety of models that will yield a specified output. But, by carrying the matching process down to the level of the individual decision processes the tests become more and more discriminating. In brief, the strength of the test can be determined by the experimenter; and our confidence in the empirical validity of the model is manifestly a function of the level of detail at which this testing procedure is carried out.

For example, in order to determine the trust investment model's ability to reproduce the portfolio selection process of the trust investor, one set of tests consisted of comparing against each other four specific portfolios chosen both by the model and the investor for the same accounts during the first and third quarters of 1960. To achieve a perfect score, the model not only had to select the correct number of securities for each portfolio, but it also had to choose the same stocks and the same number of shares of each security as was purchased by the trust investor. As can be seen from the two examples in Figure 1, the similarity between the two sets of portfolios is quite striking.^{12/}

Since there are no doubt a variety of models that could generate the same portfolios, the next series of tests are concerned with determining

^{12/}For a detailed presentation and analysis of these portfolios see: G.P.E. Clarkson, Portfolio Selection, op. cit., Chapter 6.

whether the model's decision processes are consistent with the trust investor's. To conduct this test a record of the model's decision behavior was made which was then compared to the statements recorded in the investor's protocols. For the model to pass these tests its behavior has to be sufficiently similar to the trust investor's so that a close inspection of the two streams of behavior do not provide a basis for deciding which is produced by the investor and which by the model. While these tests did not unequivocally confirm the model as well as its individual decision processes, the evidence is such that it supports the hypothesis that the model, and hence the theory, is sufficient to account for a considerable portion of the trust investment process.^{13/}

It is apparent, from this discussion, that theories of decision behavior can be subjected to a series of empirical tests. Moreover, these tests can be applied to the theory as a whole as well as to the theory's individual hypotheses. As a result, Turing's Test is a powerful method for determining the empirical validity of theories whose object is to explain human decision-making behavior.

3. The Problem of Errors

Unfortunately, the discriminatory power of these tests is somewhat impaired by the absence of suitable measures for assessing the "type" and "degree" of a model's failure. That is to say, although some models may account for observed behavior with great accuracy, others will not be so successful. Hence, the question immediately arises of how to classify, identify and measure the

^{13/} The evidence and the tests are presented in detail in: Ibid., Chapter 7.

FIGURE 1

Account 1, Selected January 8, 1960

Investment Policy: High growth with little concern for dividend income,
total funds: \$22,000

<u>MODEL'S PORTFOLIO</u>				<u>INVESTOR'S PORTFOLIO</u>			
<u>Shares</u>	<u>Stock</u>	<u>Price</u>	<u>Total</u>	<u>Shares</u>	<u>Stock</u>	<u>Price</u>	<u>Total</u>
60	General American Trans- port Company	\$ 70	4,200	30	Corning Glass	\$145	4,350
50	Dow Chemical	99	4,950	50	Dow Chemical	98	4,900
10	IBM	440	4,400	10	IBM	440	4,400
60	Merck and Company	79	4,740	50	Merck and Company	80	4,000
45	Owens Corning Fiberglas	88	3,960	50	Owens Corning Fiberglas	89	4,450
	<u>ESTIMATED Yield 1.6%</u>		<u>\$22,250</u>		<u>ESTIMATED Yield 1.6%</u>		<u>\$22,100</u>

Account 2, Selected June 10, 1960

Investment Policy: High income with possibility of price appreciation,
total funds: \$37,500

<u>MODEL'S PORTFOLIO</u>				<u>INVESTOR'S PORTFOLIO</u>			
<u>Shares</u>		<u>Price</u>	<u>Total</u>	<u>Shares</u>		<u>Price</u>	<u>Total</u>
100	American Can Company	37	\$3,700	100	American Can Co.	38	\$3,800
100	Continental Insurance	51	5,100	100	Continental Ins.	51	5,100
100	Equitable Gas Company	36	3,600	100	Equitable Gas Co.	35	3,500
100	Duguesne Light Company	24	2,400	100	General Public Utility	24	2,400
100	Libbey Owens Ford	50	5,000	100	Libbey Ownes Ford	51	5,100
100	International Harvester	44	4,400	50	National Lead	91	4,550
100	Philadelphia Electric	49	4,900	100	Philadelphia Elec.	49	4,900
100	Phillips Petroleum	43	4,300	100	Phillips Petroleum	43	4,300
100	Socony Mobil	37	3,700	100	Socony Mobil	36	3,600
	<u>ESTIMATED Yield 4.8%</u>		<u>\$37,100</u>		<u>ESTIMATES Yield 4.6%</u>		<u>\$37,250</u>

types of "errors" such models must contain. One answer is to postulate that all errors are due to an incorrect specification of the model's decision processes. If this rule is taken as the principal criterion then whenever errors occur this is a signal to go back and retest the appropriate parts of the model until such time as its output can account for the observed.^{14/}

Such a rule, despite its apparent simplicity, does not provide a complete answer to the problem. If all errors in the model's behavior are considered as errors in its decision processes then theorists will be motivated to include as many hypotheses and parameters as are necessary to produce the desired stream of behavior. Consequently, models will tend to contain an abundance of free parameters and general rules about parsimony will tend to be ignored. This is not to say that as these models are subjected to an increasing number of empirical tests, excess parameters will not be deleted where possible. Rather, it is being suggested that unless some measures are developed which permit the determination of the degree to which a model fails a particular test, the tendency will be to construct models which have a large number of free parameters and as a result are capable of being "fitted" to a wide range of observed behavior. The problem is clearly one of how to distinguish between models which are in some reasonable sense empirically "true" from those which are corroborated simply because they contain so many free parameters that they can be fitted to the available data. Turing's Test is a method for discriminating

^{14/} For a further discussion of the error problem see: Part 2, "Simulation of Cognitive Processes" in E.A. Fergenbaum and J. Feldman, op. cit.

between those models which can and those which cannot produce behavior that is indistinguishable from its human counterpart. But of the set of models which pass this test, how is one to avoid accepting as empirically confirmed models which have passed the test for essentially trivial reasons? In effect, the answer to this question is no different for models of decision processes than it is for other models in empirical science. In the physical sciences it is never possible to tell whether a particular theory or its model is empirically true. The best that can ever be said is that so far it has not been disconfirmed by all of the tests to which it has been submitted. Thus, until such time as a theory is disconfirmed or replaced by one which is more comprehensive, the theory must be accepted as it stands--an empirically testable theory of a particular set of behavior.

To illustrate these remarks consider for a moment the task of deciding whether a specific model of the theory of problem solving encompassed by the General Problem Solver is to be considered corroborated or confuted by a particular test. The model in question was constructed to account for the behavior of certain subjects when they were engaged in the solution of a set of problems in symbolic logic. After providing the model with the requisite vocabulary, a set of definitions sufficient to allow it to consider problems in symbolic logic, and the initial conditions, i.e. the axioms of the system and the theorems to be proved, the model was instructed to develop the required proofs.^{15/} Concurrently, protocols were taken of the decision processes of a number of

^{15/}See Chapter 12, p. 246-47 and the references for the principal hypotheses of GPS.

students. Their task being to construct proofs for the same theorems. To test the model the output of the model's processes is compared with the recorded behavior of the subjects. Such a comparison is provided by the following excerpts from the decision behavior of a student and the model, when they were considering the problem of transforming the statement $\underline{R} \cdot (\sim \underline{P} \supset \underline{Q})$ into the statement $(\underline{Q} \ \underline{P}) \cdot \underline{R}$.^{16/}

Model's Behavior

Student's Protocol

- | | |
|--|---|
| 1. $\underline{L}_0$: $(\underline{Q} \ \underline{P}) \cdot \underline{R}$ | (Expression to be obtained) ^{17/} |
| 2. $\underline{L}_1$: $\underline{R} \cdot (\sim \underline{P} \supset \underline{Q})$ | (Expression given at start) |
| 3. Goal 0: Transform $\underline{L}_1$ and $\underline{L}_0$ | (Goal set by experimenter) |
| 4. Match gives position difference
($\underline{\Delta} \ p$) | I'm looking at the idea of reversing these two things now. |
| 5. Goal 1: Reduce $\underline{\Delta} \ p$ between $\underline{L}_1$ and $\underline{L}_0$ | (Thinking about reversing what?) |
| 6. Search list of rules | The $\underline{R}$'s..... |
| 7. Goal 2: Apply $\underline{R}_1$ to $\underline{L}_1$ | |
| 8. Match: $\underline{R}_1$ applicable | |
| 9. Test rule functions:
reduces $\underline{\Delta} \ p$ | Then I'd have a similar group at the beginning but that seems to be.. |
| 10. no others | |
| 11. <u>Set to execute</u> $\underline{R}_1$ when analysis complete | I could easily leave something like that to the end, then I'll..... |

* * *

* * *

^{16/} For a detailed discussion of the model and this test see: A. Newell and H.A. Simon, "The Simulation of Human Thought," op. cit., pp. 155-176.

^{17/} Statements in parentheses are experimenter's statements and explanatory statements. All other statements are the subject's.

Model's Behavior

Student's Protocol

23. Goal 7: Apply R_3 to right L_1

Well....then I look down at Rule 3

24. Match: R_3 not applicable

and that doesn't look any too practical

25. Goal 8: Apply R_4 to L_1

Now 4 looks interesting

26. Match: R_4 not applicable^{18/}

Its got three parts similar to that...and...there are dots so that connective...seems to work easily enough,

28. Test rule functions:
doesn't reduce $\triangle p$

but there's no switching of order.

* * *

* * *

33. Search rules again, but don't

I need that P and Q changes so...

34. reject without attacking subproblem

35. Goal 14: Apply R_1 to right L_1

36. Match: R_1 fails, right
right L_1 has $\supset (\triangle c)$

I've got a horseshoe there.

37. Test rule functions: reduces $\triangle P$,
no others

That doesn't seem practical any place through here.

38. Set to execute R_1 , if applicable

39. Goal 15: Reduce $\triangle c$ between right
 L_1 and R_1

I'm looking for a way now, to get rid of that horseshoe.

40. Search list of rules, for rule

41. with $\supset$ that reduces $\triangle c$

^{18/}But the subject mistakenly thinks R_4 is applicable; therefore tests its function.

Model's Behavior

Student's Protocol

42. Goal 16: Apply R_6 to right L_1

Ah...here it is, Rule 6.

43. Match: R_6 applicable

* * *

* * *

67. Match: L_4 identical with L_0

And...that's it.^{19/}

Given this evidence is it possible to decide whether the model has been confirmed or disconfirmed by this test? Manifestly, this is a difficult question to answer directly. For in a number of ways the decision processes of the model closely parallel those of the student. Yet there are cases, notably line 26, where the student errs and unnecessarily proceeds to test the function. Also the model examines the applicability of all the rules, while the student provides evidence of only examining the first few. This is not to say that the model's decision behavior does not come close to matching the student's, especially if one imagines the model generating grammatical sentences instead of chopped up statements. But since a method has not yet been devised for measuring the difference between these two streams of verbal behavior it is not possible in this case to directly answer the question: how close is close enough? If a technique were available for measuring the difference between two sets of verbal behavior, then it would be possible to inspect these models for excess parameters. For, if the "goodness of performance" is measurable, empirical

^{19/} A complete listing of these statements from which these are taken is given in Ibid., pp. 171-173.

explorations of the model will permit the identification of those parameters and processes which can be deleted without lowering the level of performance below some acceptable standard. Such is the case, if the model produces an output which is amenable to numerical analysis.

For example, if a theory is concerned with the pricing process within a firm, part of a particular model's output will be a collection of items to each of which is attached a specific price. The actual prices set for the items can be readily observed and the differences between these two sets of prices noted. For a given level of predictive success, say 90%, the model can now be experimented with to see which set of processes and parameters can be deleted so that its ability to predict the actual prices never falls below 90%.^{20/} Consequently, as long as it is possible to measure a model's predictive success, and as long as it is possible to agree on the significance of specific levels of success, then empirical explorations can be conducted so that excess parameters and processes are deleted and parsimony preserved.

The essence of this testing procedure is to employ predictions to confute or corroborate a theory's hypotheses. One striking way to accomplish same would be to infer from an existing model one or more propositions about decision behavior which have not yet been put to test. If they are not disconfirmed then

^{20/} It should be noted that a theory has already been constructed to account for pricing decisions in a department store. Further, in the tests that have been conducted, a particular model predicts correct prices, including special sale prices and mark-downs, approximately 95% of the time. While tests have not been conducted to determine the number, if any, of excess parameters and processes, this model meets the conditions required for such empirical explorations. The model is described and the data are presented in: R.M. Cyert and J.G. March, A Behavioral Theory of the Firm, Prentice-Hall, Englewood Cliffs, 1963, Chapter 7.

such evidence would provide strong support for the original hypotheses. Such tests can serve as corroborative evidence because the model's hypotheses can also be subjected to test. If this were not the case, then theories of decision behavior would no more be a part of empirical science than those of econometrics or classical economics.

4. Heuristics, Algorithms and Statistical Tests

So far the discussion has focused on the problems raised by subjecting theories of decision processes to Turing's Test. While it has been noted that this test can be applied in varying strengths it has also been pointed out that when comparing two sets of verbal behavior it is not yet possible to reliably measure the degree of difference or error between them. Despite the absence of such a measure it is still possible to distinguish between models which are sufficient to account for observed behavior and those that are not. Accordingly, even though this difference is not expressible as a numerical function of specific variables, these models can be subjected to a series of empirical tests.

It is important to remember that the testing procedure can be applied to the model's decision processes and that the limit of detail at which this testing can take place is defined by the level of detail recorded in the protocols. By emphasizing the fact that the decision processes can be submitted to test a criterion is provided for identifying the types of decision processes that are appropriate for any specific model. In particular, it provides a criterion by which the appropriateness of heuristic or algorithmic decision rules can be decided.

To illustrate this point consider for a moment the decision processes which are employed to select portfolios in the investment model. These processes

are stated in the form of heuristics--i.e., they describe the search and selection procedures which delimit the available alternatives. The restricted set of alternatives in turn provides the basis for the final selection. A selection is made by choosing the first security that passes the relevant criteria from each of a number of industry lists. Thus, selections are made sequentially, and many securities, although quite suited for the particular portfolio, may never even be brought up for consideration. On the other hand, if a set of algorithmic decision rules were employed, e.g. choose only those securities which optimize some criterion function where the process of optimization was specified in complete detail, the behavior generated by such processes would be quite different from that of the investment model. Since the decision behavior generated by both these types of decision rules can be compared with behavior recorded in protocols, it is clear that on this basis it is possible to reject those processes that are inconsistent with the observed behavior.

For example, assume for the moment that one wants to test the hypothesis that the trust investor employs algorithmic decision rules. Further, assume that these rules take the form of some optimizing routine, e.g. choose those securities which subject to certain constraints maximize expected returns.^{21/} With such a decision rule it is clear that the model will examine all of the relevant alternatives before selecting its portfolio--an optimizing decision rule implies that all alternatives are to be examined before a choice is made. Accordingly,

^{21/} For an excellent exposition of such a position see: H.A. Latane, "The Rationality Model in Organizational Decision-Making," in H.J. Leavitt, (ed), The Social Science of Organization, Prentice-Hall, Inc., Englewood Cliffs, 1963, pp. 85-136.

this pattern of search and selection behavior can be contrasted with the observed, and consistencies as well as inconsistencies noted. Also, if an algorithm of this sort is employed, processing time should be roughly equivalent for each security examined. For, given the decision rules and the appropriate data, there is no reason to suppose it will take longer to evaluate the expected returns of one security rather than another. But human processing time can also be observed. And if, as was the case in the investment study, some securities are accepted or rejected quite rapidly while others are processed for longer periods, then this evidence would tend to confute what might be called the algorithmic hypothesis. When the decision processes themselves are subjected to a process of refutation by empirical test, the debate over heuristic vs. algorithmic decision rules rapidly disappears. For it is highly unlikely that two such different types of processes could generate identical streams of decision behavior. The problem of choice is simply resolved by selecting the one that most closely accounts for observed data.^{22/}

When carrying out such a series of tests it has already been mentioned that if the model's predictive success can be adequately measured, then excess free parameters can be isolated and deleted by repeated empirical explorations. The point to note in this respect is that statistical tests are unfortunately of slight value in helping to isolate the surplus or excess free parameters. If these models were stated in terms of standard difference, differential, or stochastic

^{22/} For a stimulating example of the power of a specific set of heuristic decision rules to reproduce the search and selection procedures of grandmaster and expert chess players see: H.A. Simon and P.A. Simon, "Trial and Error Search in Solving Difficult Problems: Evidence from the Game of Chess," Behavioral Science, Vol. 7, October, 1962, pp. 425-429.

equations with a limited number of independent variables, then the problem of excess parameters could be answered within the confines of the identification problem.^{23/} But decision theories are stated in terms of programs of processing rules which are not amenable to a similar mathematical analysis. Despite the fact that the problem could be considered in an analogous manner the standard mathematical methods of solution are no longer applicable.

For example, consider the problem of estimating the statistical significance of each parameter in a small set of decision processes. Assume for the moment that these processes describe a sequential selection procedure which consists of fifteen different items. Also assume that a statistical test is to be employed, say an analysis of variance test, which will permit the delineation of those parameters that play a statistically significant role in accounting for the observed data. In order to conduct this test two items must be known or be capable of being estimated: the density function of the population from which the data is to be drawn, and the sample size required for statistically significant results to be produced.

Taking these points in reverse order it is clear that if samples of twenty will generate significant results, and if each of the fifteen items were employed each time the decision process were used, then twenty experiments would provide the requisite data. But it frequently happens that not all decision points are evoked each time a decision process is employed. If it takes on the average twenty-five applications of the selection process to ensure that each item has been

^{23/} See Chapter 8, section 3 for a discussion of this subject.

evoked at least once, then it will take approximately five hundred experiments to generate samples of at least twenty observations for each parameter. Further, if the selection process itself is not employed each time this part of the model is subjected to empirical test (e.g. if the process in question is an infrequently used sub-routine), the number of experiments required has now increased to a very large and impractical number.^{24/} This, however, is a practical difficulty which can no doubt in many cases be met by practical expedients. The principal obstacle lies in estimating the characteristics of the population from which the data are drawn. In an artificial case the data can be selected from nearly normalized populations.^{25/} But when dealing with theories that are intended to account for observed behavior the data are provided by environment. And as noted above, the empirical determination of population density functions is not without its difficulties.^{26/}

From this brief example it is clear that it is simple not feasible to employ standard techniques to examine the sample distributions and to evaluate the statistical significance of each item in a theory of decision behavior. This is not to suggest that the task of eliminating excess parameters is, for all practical purposes, hopeless. Rather, the problem must be approached in the same

^{24/} For an excellent example of the problems posed by the measurement of the statistical significance of specific parameters in a behaving system see: C.P. Bonini, Simulation of Information and Decision Systems in the Firm, Prentice-Hall, Englewood Cliffs, 1963, Chapters 7 and 8.

^{25/} See for example: M.A. Geisler, "The Sizes of Simulation Samples Required to Compute Certain Inventory Characteristics with Stated Precision and Confidence," Management Science, Vol. 10, January, 1964, pp. 261-286.

^{26/} See the discussions of this point in Chapters 9 and 10, sections 2, and 2 and 3 respectively.

spirit as is the development and testing of theories in any branch of science. There theoretical speculations are controlled and refined by direct confrontation with empirical observation. Therefore, although the primary goal may be to simplify and increase the power of our theories, progress can only be achieved by the diligent application of empirical test.

Chapter 14

THEORIES OF ECONOMIC BEHAVIOR - THE CONSUMER AND THE FIRM

In the previous two chapters the discussion is directed toward an examination of the task of developing theories of individual decision behavior. While some of the evidence cited both in the text and the references is concerned with economic decision processes, it is clear from the remainder that the empirical content of these theories of decision-making behavior resides in the ability to test their hypotheses against a diversity of observable behavior. Our primary interest as economists, however, is to be able to explain and predict the behavior of aggregates of individuals and not just the behavior of the individuals themselves. This is not to say that theories of individual behavior are of little concern to economists. Rather, our interests will be best served if it is possible to develop testable theories with which the behavior of collections of individuals, such as consumers, as well as groups or organizations of individuals, such as firms can be explained. Moreover, to fully analyze the behavior of an economy it is also necessary to be able to explain the interactions among individuals, and among individuals and firms such as takes place in various markets. In order to develop such theories it is clear that the first task is to isolate and identify a set of testable relations which can serve as the empirical base for these theoretical structures.

One method of approach is to start with a theory of organizational or firm behavior from which it is possible to deduce testable hypotheses of decision behavior. If the inferred relations are in fact corroborated, then they would constitute the beginnings of scientific theory of organizational or firm decision-making behavior. Manifestly this is, in outline form, the general procedure which guided the development of the classical theory of the firm. It is also apparent from the analysis in Part II that the classic approach is not able to produce the desired, testable hypotheses. However, the existence of a behavioral theory of the firm^{1/} -- a theory of firm behavior based upon observations of organizational decision processes -- lends support to this general strategy. If it is to succeed, it must be possible to deduce the required empirical relations as well as demonstrate that they are able to survive repeated tests. Yet even if by employing this theory it were possible to establish a set of empirical relations about firm behavior, only one part of the total task would have been accomplished. This is not to suggest that to have constructed a testable theory of firm behavior would not be an important accomplishment. But it would only enable one to account for one of the three classes of behavior noted in the preceding paragraph.

^{1/}R. M. Cyert and J. G. March, A Behavioral Theory of the Firm, op. cit.

A second, a ideally more complete, solution can perhaps be found by beginning with theories of individual, economic decision-making behavior. Such theories must be capable of scoring repeated empirical tests, and, as already noted, are to be based on a wide variety of observable decision behavior. The second step is to employ the theories of individual behavior as the empirical basis for theories of organizational or firm behavior. In effect, I am proposing that a resolution of the theoretical and empirical obstacles can be found by "reducing"^{2/} existing theories of organizational and firm behavior to testable theories of individual decision behavior.

For this reduction process to succeed it implies that the theories of individual, economic decision behavior must be constructed in such a fashion that they are sufficient to account for individual as well as group behavior. In order for this result to occur two conditions must be satisfied. The first is that the laws or hypotheses of group or organizational theories must be deducible from the theories of individual behavior. If a theory of organizational or

^{2/}For an excellent discussion of the process of reduction in empirical science see: E. Nagel, The Structure of Science, Harcourt, Brace and World, New York, 1961, Ch. 11; and P. Oppenheim and H. Putnam, "Unity of Science as a Working Hypothesis," in H. Feigl, et al. (eds.), Minnesota Studies in the Philosophy of Science, University of Minnesota Press, Vol. II, 1958, pp. 3-36.

This argument, with respect to the development of a testable theory of consumer behavior, is presented in greater detail in: G. P. E. Clarkson, The Theory of Consumer Demand, op. cit., Ch. 7.

firm behavior contains terms and expressions which do not appear in the relevant theory of individual behavior then it is not possible to immediately meet the first criterion. In this case various assumptions or further hypotheses must be introduced to link the terms in the theory of individual behavior to the terms and relations contained in the organizational theory. For example, if hypotheses about the role of goals and the resolution of conflict in the structure of organizational decision processes are to be inferred, then the theory accounting for individual behavior must either already contain these terms and expressions or further postulates must be introduced to permit the derivation to take place.

The second main condition is that the basic postulates or hypotheses of the individual theory must be both testable and reasonably well confirmed^{3/} by the available evidence. The purpose of this criterion is to ensure that essentially trivial reduction theories are not constructed. It would not be an important scientific accomplishment merely to develop a set of hypotheses about individual behavior from which theories of firm and organizational behavior could be deduced, if these theories could not be subjected to empirical test. Hence, before a theory of individual behavior can be accepted as a possible basis for this

^{3/}The condition of being "reasonably well confirmed" is admittedly vague. But the problems involved in determining a degree of confirmation go beyond the scope of our analysis. For a detailed and lucid presentation of one interpretation of the meaning of "degree of confirmation" see: R. Carnap, Logical Foundation of Probability, University of Chicago Press, Chicago, 1950, and "Statistical and Inductive Probability," in E. H. Madded, (ed.), The Structure of Scientific Thought. Houghton Mifflin, Boston, 1960, pp. 269-279. For a different interpretation see: K. R. Popper, Conjectures and Refutations, Reutledge and Kegan Paul, London, 1963, especially Chs. 3, 10 and 11.

scientific venture it must be demonstrated that its hypotheses are both capable of test and have already survived a number of such tests.

In the preceding two chapters a theory is presented which meets these formal requirements--i.e., the theory contains testable hypotheses and the available evidence demonstrates that some of these hypotheses have survived a number of tests. Since the formal criteria are satisfied the task that has yet to be completed is to develop the economic theories of firms, consumers and markets that are consistent with this approach. Accordingly, the remainder of this chapter and the whole of the next are devoted to an examination of the methods by which the required theories can be developed.

1. On a Theory of Consumer Behavior^{4/}

In order to explain the decision behavior of consumers a theory is needed which can be employed to account for the observable diversity of decisions made by individual consumers. While some economists may be particularly interested in consumer decisions with respect to purchases of durables, others are interested in the processes which determine the purchases of comestibles, clothing, entertainment, etc. Concurrently, there are still other investigators who are concerned

^{4/}This section is indebted to: G. P. E. Clarkson, The Theory of Consumer Demand: A Critical Appraisal, op. cit., Ch. 8.

with the processes by which specific types, makes or brands of articles are purchased within a given category. If a theory is to explain consumer behavior, it must be possible to adapt it to each of these varying circumstances. That is to say, it must be constructed in such a fashion so that with the addition of the appropriate information and interpretive rules it can be applied with equal success to each of these specific decision situations.

To develop such a theory there are two possible strategies which might be adopted. The first would be to inspect in detail the decision processes of a number of consumers for each of the major commodity categories. From such observations theories of these specific decision processes would be constructed so that the observed behavior of these consumers could be explained. Once the theories had survived a number of tests the theories themselves would be examined in order to detect the general characteristics that they had in common. From such general characteristics a general theory of consumer behavior would evolve which in turn, with suitable amendments, could be applied to explain specific sets of behavior.

A second method of approach would be to begin with the general postulate of the invariance of the structure of decision processes among decision-makers. This postulate states that the structure of decision processes is the same for all decision-makers. Thus, the structure of a general theory of consumer behavior can be directly inferred from the structure of the theory of individual decision behavior. A general structure, however, cannot serve directly as a testable

theory of a particular set of behavior. In order to function in this manner the general structure has to be conjoined with the appropriate detail on the relevant decision processes. In effect, this would entail adding to the general structure the requisite detailed processes and their concomitant data, the latter being developed by an empirical analysis of consumer decision behavior.

Both of these strategies clearly require a detailed study of consumer decision behavior. At the same time, they are based on the general premise that the object of this theoretical exercise should be a general theory of consumer behavior which can be adapted to the explanation of specific events by the inclusion of certain data and decision rules. That is to say, if a general theory of consumer behavior is to be developed it can be considered as an "ideal" theory, where the additional data and decision processes are the interpretive rules which permit the ideal theory to be related to and to explain the behavior of a particular consumer.^{5/}

To illustrate this general approach consider once again the general theory of human problem solving, GPS. This theory essentially consists of two separate components. The first is a set of general hypotheses about the structure and content of problem solving decision processes. These hypotheses include such items as general methods for solving problems, the basic decision processes

^{5/} It should be noted that the natural sciences frequently use this technique of formulating theories to account for "ideal" cases. Almost all the well-known physical laws, e.g., the gas and gravitational theories, are formulated in this manner; and as long as a set of interpretive rules exist these theories can be tested against actual observations. See, for example, the excellent discussions of this point in: E. Nagel, "Problems of Concept and Theory Formation in the Social Sciences," op. cit.; and C.G. Hempel, "Typological Methods in the Social Sciences," op. cit. For further examples see: J.W.N. Watkins, "Ideal Types and Historical Explanation," in H. Feigl and M. Brodbeck, op. cit., pp. 723-743.

available to the theory and the structure of the memory. For the theory to be applied to a specific situation the data and decision rules pertinent to the second component must be added. These items consist of the vocabulary, special definitions, and other data necessary to interpret the specific problem situation for the theory.^{6/} As a result, GPS is a general or ideal theory of human problem solving which must be conjoined with the appropriate interpretive rules before it can be employed to explain the decision behavior of an individual problem solver.

In a similar manner, therefore, a theory of consumer behavior would consist of general hypotheses about consumer decision processes, basic informative processes, and the structure of memory which when appropriately interpreted would be sufficient to explain observable behavior. To develop such a theory it is clearly necessary to construct both components, i.e. the ideal theory, and the interpretive rules. The former can in part be derived from the theory of individual behavior. But the remainder of the general theory as well as the interpretive rules can only be developed from a detailed inspection of consumer behavior. While a general theory could apply to the behavior of one or more consumers, to be able to test the theory the interpretive rules must specify whether it is the behavior of a single consumer or groups of consumers that is to be explained. To test a theory specific data must be employed. Accordingly, if the behavior of groups of consumers is to be explained the interpretive rules

^{6/} See Chapter 12, pp. 246-247.

must pertain to the decision processes of such aggregates. Similarly, if it is an individual's behavior that is under investigation, the interpretive rules need only pertain to this specific consumer. Hence, as already noted, although the structure of the general theory can be derived from the theory of decision behavior, it is only by empirical exploration of consumer behavior that the detailed specification of this structure and the interpretive rules can be determined.

To illustrate these remarks consider the general characteristics of consumer behavior that would need to be included if the theory were to account for observable decision behavior. Since each consumer has a certain level of income a decision process is required which will allocate this income over the various classes of commodities. While consumers may differ in the proportions of their income which they allocate to each set of commodities, this allocative process is common to all consumers. Further, within any particular social and economic stratum regularities may appear among the specific proportions selected by these consumers. Such regularities, if they are confirmed by empirical research, can also be inspected for the rate at which they change over time. If, as the evidence appears to indicate,^{7/} the allocative process is reasonably stable these results would immediately lead to a specification of one part of the allocative decision process. Such a process is summarized by

^{7/} Much of the evidence from recent research on consumer behavior is reviewed in R. Ferber, "Research on Household Behavior," American Economic Review, Vol. 52, March 1962, pp. 19-63.

the following three postulates: (i) Each consumer decides over a given interval of time on the proportion of his total income to be spent on each category of commodities;^{8/} (ii) this decision procedure remains constant over time, as long as total income does not vary significantly; and (iii) the proportions of total income a consumer allocates to each category are closely approximated by the proportions allocated to these same commodity classes by those consumers who within a given geographic location are in the same social and economic position.

With these three postulates it is clearly possible to begin to observe the allocative decision procedures of individuals as well as groups of consumers. If the theory is to account for an individual's allocations the actual proportions employed by this individual must be observed and entered as specific parameter values. Similarly, if the theory is to explain a group's allocative procedure only the group's proportions need be observed. Accordingly, given these three postulates, and assuming for the moment that they are supported by empirical test, it is clear that the observed proportions constitute the requisite interpretive rules.

Once a consumer chooses to spend a certain proportion of his income on a particular commodity category, say food, he is then faced with the problem of deciding how to allocate these funds among the possible types of comestibles.

^{8/}It should be noted that the notion of such an allocative process is not a novel idea. For a theoretical formulation within classic utility analysis see: R.H. Strotz, "The Empirical Implications of a Utility Tree," Econometrica, Vol. 25, 1957, pp. 269-280; and I.F. Pearce, "An Exact Method of Consumer Demand Analysis," Econometrica, Vol. 29, October, 1961, pp. 499-516.

To understand this decision process, it is necessary to examine the decision procedures that govern the expenditure of these funds. Similarly, if the theory is to account for all expenditures, then the decision processes relevant to each commodity category must also be determined. Such a theory would be somewhat large and complex. Moreover, if all processes have to be empirically checked against the behavior of each consumer, the task of constructing the desired body of theory would be very demanding. However, if there are some similarities among the decision processes relevant to each commodity category, then the problem of developing a theory of consumer behavior may not be as formidable as previously expected.

For example, as noted above, the General Problem Solver principally consists of a set of hypotheses which describe the processes by which humans solve certain types of problems. These hypotheses contain no references to the subject matter of any specific problem. Consequently, when the theory is employed to account for the behavior of individuals proving theorems in symbolic logic the theory has to be interpreted by providing it with the requisite vocabulary, axioms, and rules of inference. Similar interpretive rules must be provided if the theory is to account for the behavior of subjects proving theorems in geometry, deciding on moves in chess, or tackling other problems which are consistent with the means-ends analysis of the theory. As a result, if a single theory is to encompass the decision processes of individuals as well as groups of consumers, and if at the same time it is to reflect the principal characteristics of the general theory of human problem solving, it is evident that a substantial proportion of its hypotheses are to be stated in such a way that they are independent of the particulars relevant to a specific commodity

category. This implies that the decisions processes sufficient to account for, say, the allocative decisions within one class of commodities are also sufficient to account for the allocation of funds within any of the remaining categories. In brief, such a theory assumes that consumers employ largely similar sets of decision processes to solve all of their allocation and purchasing decisions.

A. Some Possible Processes and Interpretive Rules

In order to guide the development of some of the principal decision processes as well as their respective interpretive rules the basic requirements of a theory of consumer behavior can be stated as follows: (i) the principal decision processes are to consist of a single set which can be applied to the allocation of funds among commodity categories as well as to the selection of individual items within any specific category; (ii) these decision processes are to be constructed so that they are independent of the subject matter of any one class of commodities; and (iii) for each category of commodities some specific decision processes are required so that the processes in (i) can be applied to the particular decisions that occur within each of the individual categories. Given these requirements, it is now possible to examine some of the decision processes that, subject to corroboration by actual investigations, could be constructed to be largely independent of the particular contents of any single class of commodities. Concurrently, once these processes are specified, it is then possible to note the interpretive rules that must be provided if the resulting theory is to account for the observed behavior of individuals or collections of individual consumers.

The first decision procedure that could be constructed in this manner is the process by which a consumer decides how to pay for a particular purchase. While, at first sight, this may not appear to be a particularly important process, its function would be to determine whether the item under consideration is to be paid for by cash or cash equivalents, or by a set of monthly payments. A general process of this sort could be constructed to encompass such decisions as:

(i) Whether to rent or purchase housing accommodation. For if the decision is to purchase a house, the decision process would include the size of the mortgage, interest and tax payments that could be afforded. As a result, it would also include the decision on the price that a consumer would be willing to pay for his housing. (ii) Whether to purchase other durables for cash or by accepting credit to spread the payments over a period of time. Since a separate process allocates the total funds to the separate categories, this decision process would also include a mechanism for specifying the upper limit of these periodic payments for each category. (iii) The remaining rent or buy decision that a consumer has to make from time to time.

Even though the explanation of each of these decisions may require a separate process, they all have certain elements in common--namely, whether there already are allocated funds available to cover the intended purchase or, if not, whether by accepting credit the periodic payments are low enough to permit them to be paid for out of the available funds for that class of commodities. Manifestly, this process can be constructed so that it is independent of the commodity category, and where its object is to determine within the amount of funds allocated to each category how a particular purchase is to be financed. While this second allocative process may differ in detail among individuals it would be a postulate

of this theory that its principal components could be represented by a single set of decision processes.

In order to subject such a process to empirical test it is necessary to provide some interpretive rules. Since the "buy-now-pay-later" decision rule is dependent on the process which allocates funds to the separate categories, both processes must be given an empirical interpretation before tests can be conducted. For the first decision process the interpretive rules, as noted above, are readily identifiable. All that is needed is to observe the proportions of total income a consumer or group of consumers allocate to each commodity category. Clearly, this process would account for the observed allocations up until that moment when the proportions were altered. In order to accommodate such observable changes a set of adaptive mechanisms would have to be included which would allow these proportions to change as total income rose or fell.^{9/} Once this process is empirically determined it is then necessary to examine this second allocative procedure--i.e., the process by which the funds per category are spent. The object, of course, is to empirically determine the parameters, e.g., the interest rate, the amount of credit already outstanding, the cost of the item, etc., as well as the specific processes which are sufficient to account for this part of the consumer's decision process.

^{9/}Two such adaptive mechanisms are suggested by the "permanent income hypotheses" of M. Friedman, A Theory of the Consumption Function, National Bureau of Economic Research, Princeton, 1957, and the "permanent wealth," hypothesis of F. Modigliani and A. Ando, "The 'Permanent Income' and the 'Life Cycle' Hypotheses of Saving Behavior," in I. Friend and R. Jones (eds.), Proceedings of the Conference on Consumption and Saving, University of Pennsylvania, Vol. II, 1960, pp. 49-174.

of this theory that its principal components could be represented by a single set of decision processes.

In order to subject such a process to empirical test it is necessary to provide some interpretive rules. Since the "buy-now-pay-later" decision rule is dependent on the process which allocates funds to the separate categories, both processes must be given an empirical interpretation before tests can be conducted. For the first decision process the interpretive rules, as noted above, are readily identifiable. All that is needed is to observe the proportions of total income a consumer or group of consumers allocate to each commodity category. Clearly, this process would account for the observed allocations up until that moment when the proportions were altered. In order to accommodate such observable changes a set of adaptive mechanisms would have to be included which would allow these proportions to change as total income rose or fell.^{9/} Once this process is empirically determined it is then necessary to examine this second allocative procedure--i.e., the process by which the funds per category are spent. The object, of course, is to empirically determine the parameters, e.g., the interest rate, the amount of credit already outstanding, the cost of the item, etc., as well as the specific processes which are sufficient to account for this part of the consumer's decision process.

^{9/} Two such adaptive mechanisms are suggested by the "permanent income hypotheses" of M. Friedman, A Theory of the Consumption Function, National Bureau of Economic Research, Princeton, 1957, and the "permanent wealth," hypothesis of F. Modigliani and A. Ando, "The 'Permanent Income' and the 'Life Cycle' Hypotheses of Saving Behavior," in I. Friend and R. Jones (eds.), Proceedings of the Conference on Consumption and Saving, University of Pennsylvania, Vol. II, 1960, pp. 49-174.

In a similar manner processes could be developed which would account for the resolution of conflicts or mis-allocations of funds among the categories, the process by which expectations about future prices, product developments and other variables affect current behavior, and the actual selection procedures that permit a consumer to choose within a given category one set of commodities from those available at the time. Each of these processes would consist of a basic set of decision procedures^{10/} which would need to be empirically interpreted to account for a specific set of observed behavior.

For example, with respect to the first of these processes, the funds available for expenditure within a category at a particular time may not be sufficient to cover either the proposed purchase or the payments already incurred. One method of resolving such conflicts is by a process which prohibits further purchases in this category until further funds are allocated. Another possibility is a decision rule which permits the interchange of unspent monies between one category and another. In either event it is the task of empirical research to discover which, if either, of these processes is consistent with observed behavior.

The point to notice, however, is not whether any one or all of the decision processes outlined above do represent the actual decision procedures employed by consumers. But rather that it is possible to postulate the existence of such processes and then carry out the empirical investigations necessary to

^{10/}For further discussion of these hypothesized processes see: G.P.E. Clarkson, The Theory of Consumer Demand, op. cit., Chapter 8.

to corroborate or confute them. Consequently, although the development and specification of the requisite hypotheses and interpretive rules can only be accomplished by empirical research, the research already completed on individual decision of behavior provides a sound, empirical foundation upon which to build a testable theory of consumer behavior.

2. On a Theory of the Firm

To construct a theory of consumer behavior which would be sufficient to account for the behavior of individuals as well as groups of consumers a postulate is employed which asserts the existence of invariances in the structure of decision processes among decision-makers. While this postulate is sufficient to permit the development of theories dealing with individual behavior, a theory of organizational or firm behavior needs to account for the interactions among individuals as well as the behavior of the individuals themselves. Earlier it was pointed out that one method of developing a testable theory of firm behavior would be to reduce current theories of organizational and firm behavior to the theory of individual decision behavior. The advantage of such an approach is clear, in that the reduction process would enable the theory of individual behavior to serve as the empirical foundation for theories dealing with organizational behavior. That is to say, if such a reduction can be established then some hypotheses about organizational behavior can be tested by direct reference to individual behavior. Manifestly, many hypotheses concerning the behavior of a firm will relate to the firm's decision problems. Yet, if an empirical link can be established between the behavior of a firm and that of the individuals of which it is composed, then the empirical research on individual

behavior can be used to test, augment, and interpret hypotheses about the firm's decision making process.

To effect this reduction between existing theories of organizational and firm behavior a second postulate is required--namely, a postulate which asserts the existence of invariances between the structure of individual and organizational decision processes. The basis of this postulate resides in inductive and empirical grounds. It cannot be proved as a theorem. Indeed, the only grounds upon which it can be supported, other than by direct empirical test, is its consistency with the theory of individual decision-making behavior. Essentially, this is nothing more than an appeal to parsimony as a rule of procedure and a supposition that this is the appropriate way in which Occam's razor ought to be applied.

The empirical value of the postulate lies in the ability it provides to interpret theories of organizational behavior on the basis of the empirical theories of individual behavior. While its value to research can only be determined by empirical test, it should not be overlooked that the empirical basis for such a postulate is in part already emerging. Consider, for example, a set of hypotheses that are taken from a general theory of planning and innovation in organizations.^{11/}

- (1) "Those variables that are largely within the control of the problem-solving individual or organizational unit

^{11/} J.G. March and H.A. Simon, Organizations, Wiley, New York, 1958, pp. 170-180.

will be considered first."

- (2) "If a satisfactory program is not discovered by these means attention will be directed to changing other variables that are not under the direct control of the problem solvers."
- (3) "If a satisfactory program is still not evolved, attention will be turned to the criteria that the program must satisfy, and an effort will be made to relax these criteria so that a satisfactory program can be found."
- (4) "In the search for possible courses of action, alternatives will be tested sequentially."

Without further elaboration and specification of the empirical meaning of these variables as well as the conditions under which the hypotheses apply it would not be possible to submit them directly to empirical test. However, if the second postulate of invariance is accepted temporarily, (a postulate that is manifestly implicit in the hypotheses quoted above) these hypotheses no longer remain in an uninterpreted state. For if this postulate is employed the variables can be immediately specified and the interpretive rules determined by an empirical investigation of these hypotheses among individual decision-makers.

For instance, since a theory of individual decision behavior can be subjected to a process of refutation by empirical test, it must be possible to determine the empirical validity of the following hypotheses concerning individual behavior:

- (1) Within a given problem context individuals will select those parts of the problem to be worked on first that are within their ability to control.
- (2) If a solution cannot be reached in this manner, an individual will then direct his attention to the remaining parts of the problem that are not under his control.
- (3) If a solution is still not attained, attention will be directed to the criteria that the solution must satisfy, and an attempt will be made to relax these criteria so that a satisfactory solution can be found.
- (4) In the search for a solution, alternatives will be examined sequentially.

That these hypotheses can be subjected to test is clear. That they will in fact be corroborated can only be determined by conducting such tests.^{12/} What is more important, however, is that whether they turn out to be empirically true or false, a procedure exists for determining their empirical validity.

As a further illustration of this method of interpreting theories about organizational behavior, consider the recently proposed behavioral theory of

^{13/} Evidence strongly suggesting their empirical validity can be found in the empirical exploration of the game of chess. See, for example, the behavior described in: A. Newell, J.C. Shaw, and H.A. Simon, "Chess-Playing Programs and the Problem of Complexity," op. cit., and H.A. Simon and P.A. Simon, op. cit.

the firm.^{13/} The basic framework of this theory consists of a set of variable classes and a set of relational concepts. The chief postulate is that the firm can be represented by a set of decision-making processes, and the analysis of these decision processes is to be carried out in terms of the variable classes and the relation concepts.

Firms, under this theory, have goals as well as expectations and have to make choices among various alternatives. Goals are represented as the names of a collection of independent variables. Attached to each of these goals is an attribute known as the level-of-aspiration with respect to that goal. Aspiration levels are single valued entities which are either satisfied or unsatisfied by the operating behavior of the firm. For example, a firm could be represented as having specific goals with respect to total sales, market share, profit, production rate as well as a variety of other variables dealing with other aspects of the organization. To each of these goals would be associated a level-of-aspiration, which in the case of sales, market share, profit and production rate could be represented numerically. While the aspiration levels themselves depend on past experience, as well as on a number of other factors, at any one moment of time they are either being satisfied or not. And it is the lack of satisfaction or violation of an aspiration level which is the hypothesized mechanism by which attention is directed to the various goals.

^{13/}R.M. Cyert and J.G. March, A Behavioral Theory of the Firm, op. cit., see especially Chapter 6.

Goals are set by the firm independently of each other. During any one period of time they need neither be consistent nor compatible with one another. As a result, the organization requires some procedures whereby conflict between goals and their concomitant behavioral routines can be resolved. The behavioral theory of the firm posits that an organization divides its total decision-making process into subsections, and then assigns these separate segments to individual units within the firm. By this postulate of local rationality the total decision problem is broken down into a number of independent parts, the responsibility of each of which residing in a single, organizational subunit. In addition, the theory postulates that each decision unit employs a set of decision rules which are again based on the notion of acceptable levels of performance. Since the failure to satisfy a goal is the process by which attention is directed by the organization to a problem area, there is no reason to suppose that several goals will not require attention at any one item. In order to take care of such situations the theory postulates that the organizations will attend to these goals sequentially. Conflict between two inconsistent goals is thereby avoided. For if each problem is viewed in relative isolation, and if problems are only attended to one at a time, then conflict between two separate problems will occur infrequently since the two situations are not dealt with simultaneously.

For example, consider a firm which is encountering difficulties in meeting the fluctuating demands of its customers as well as the production problems associated with producing this varying output. If both problems were evoked and dealt with simultaneously, there would be an obvious source of conflict between those who wanted to satisfy the customers and those who wanted to smooth out the production process. But, if problems are attended to sequentially the

task of satisfying the customers will be dealt with at one time and the task of smoothing production at another. Similarly, if the problems are dealt with by separate subunits of the organization inconsistencies between the proposed solutions will go unnoticed. As a result, the inherent conflict in many situations is largely resolved by decision procedures which preclude its recognition.

Organizational expectations are included as consequence of processes which make inferences from information available at the time. These processes are represented by a number of pattern-recognition processes and simple procedures of extrapolation. Also, their behavior depends upon the way in which information is collected and processed by the firm. Accordingly, the theory includes the three postulates on information securing or search activity. The first states that a search for information is only initiated after a problem has arisen. Such search activity is classified under the general rubric of "problemistic search." In the words of the authors, "In a general way, problemistic search can be distinguished from the former because it has a goal, from the latter because it is interested in understanding only insofar as such understanding contributes to control."^{14/} As a result, it is postulated that all search activity is motivated by the existence of specific problems and is directed toward obtaining acceptable solutions to them.

^{14/} R.M. Cyert and J.G. March, ibid., p. 121.

Since search activity could be conducted in a number of ways, the next postulate states that the search for information and alternatives will begin in the neighborhood of the problem itself and in the neighborhood of the current alternative, if there is one. Essentially, this postulate argues that search activity will first be initiated by the subunit within which the problem originated. If this process is unsuccessful and if the pressure to reach a solution is sufficiently great, the search process will become more complex as different subunits of the organization enter into this decision activity. One consequence of the second postulate is that the actual search procedures used by any specific unit of the organization will be biased by the way in which this unit views the environment. For instance, if the sales department is engaged in search activity, it will view both its problem and the possible alternatives in terms of items directly connected with selling activity. Similarly a production unit will see the solution of its problems in terms of such items as workforce, production rate, inventory costs, etc. This is not to say that an adequate solution will not be found by such approaches. Rather, such behavior has been observed on a number of occasions^{15/} and forms the basis of the third postulate which states that all search activity is biased in a manner similar to that mentioned above.

The problem posed by having to choose between available alternatives, the third of the principal components of the theory, is in part accounted for by the hypotheses noted above. If goals are independent and are attended to

^{15/} See for example: D.C. Dearborn and H.A. Simon, "Selective Perception: A Note on the Departmental Identifications of Executives," Sociometry, Vol. 21, 1958, pp. 140-144.

sequentially, if acceptable-level decision rules are employed, and if alternatives are only sought when the need arises, the problem of choice is resolved down to the point where the subunit involved accepts the first one that satisfies its requirements. In other words, search is terminated and a solution is adopted as soon as one alternative appears to satisfy the goal's aspiration level.

One part of the organization's choice problem, however, concerns the way in which it contends with the element of uncertainty which is a part of all its decision-making. In this theory of organizational behavior firms are represented as trying to avoid dealing with such uncertainty. In order to do so they are hypothesized as employing decision rules which anticipate and respond only to events in the short-run. Further, by actively searching for ways of reducing uncertainty, e.g., industry trade practices, standard operating procedures, long run purchase and sale contracts, (price stabilization, etc.) a firm can rely upon its short-run decision rules as sufficient for the task at hand.

While this discussion has only noted some of the more important hypothesis it is clear, at this level of description, that the theory is stated in far too general terms to be directly submitted to empirical test. To test the theory, both the variables and concepts as well as the hypothesized decision rules must be empirically interpreted and specified in all the requisite detail.^{16/} Moreover, in order to subject the hypotheses to test large-scale experiments on, and detail observation of, a firm's decision processes are required. Although the detailed

^{16/} Two actual models of this theory, a specific price and output model and a general model of price and output, are described in ibid., Chapters 7 and 8.

model of a department store's pricing behavior evinces the practicality of conducting such tests^{17/}, the ability to test some of these hypotheses against individual behavior, provided by the second postulate of invariance, substantially reduces the experimental problem.

To illustrate these remarks consider the process by which firms are postulated to direct and control their attention and problem solving behavior. First of all, the organization is represented as having a number of independent goals. These goals are established by different subunits and to each goal there is associated an attribute called a level of aspiration. A problem situation occurs when the results of the firm's activity fail to satisfy one or more of these goals. That is to say, a problem is defined by a failure of the level of aspiration associated with a particular goal to be reached. Once this occurs a search for a solution is initiated and is only terminated when the level of aspiration is satisfied. Accordingly, one of the basic hypotheses being employed is that failure to meet some specified target (aspiration level) is the mechanism by which problems are defined and problem-solving activity is controlled. In brief, this theory posits that problem-solving activity is directly controlled by the satisfaction or the lack of satisfaction of the levels of aspiration attached to the respective goals. One consequent of this hypothesis is the statement that if all goals are at one time simultaneously satisfied then no problem-solving activity should be observed. Another is that, if more than one goal is simultaneously unsatisfied, a mechanism must be introduced which accounts for the

^{17/} See previous footnote.

order in which these problems are dealt with. The theory posits that they will be considered sequentially, but if one is to explain a particular set of behavior it will also be necessary to be able to account for the particular sequence which is observed. Now both the firm and its subunits are composed of individuals. Also it is the individuals themselves who carry out the problem-solving activity. Thus by the second postulate of invariance these two inferences can be tested by a direct examination of individual behavior. Moreover, an empirical investigation has already been undertaken to determine the process which controls the allocation of problem-solving behavior--i.e., the process that directs an individual's attention from one problem to the next.^{18/} That is to say, an experimental situation was designed to test the propositions: (1) that problem-solving will take place when an aspiration level is unsatisfied--i.e., when some goal has not been attained, and (2) that in the absence of failure there will be no problem-solving activity. This study was concerned with individual decision behavior, and aside from testing these two propositions its object was to develop a theory to explain the problem-solving control process.^{19/}

In brief, the results are quite clear and strongly negate, for individual decision behavior, both of these propositions. When all goals are satisfied subjects do not desist from problem-solving. Further, an aspiration level does not need to be violated before problem-solving begins. When a subject's behavior is

^{18/}W.F. Pounds, "A Study of Problem-Solving Control," op. cit.

^{19/}Some aspects of this study are further discussed in the next chapter.

analyzed in terms of his decision processes it appears that problem-solving is a continuous activity. Consequently, the key to the explanation of the behavior lies in discovering the processes which allocate problem-solving activity from one problem to the next. Whether this allocative process can be reconciled with a level-of-aspiration representation is not at issue here. The importance of these experimental results pertain to the development of testable theories. Indeed, these experiments demonstrate it is practicably possible to test hypotheses about organizational decision behavior by experimental investigation of individual behavior. Obviously, not all hypotheses about organizational behavior can be tested in this manner. But as long as some are suited to this approach, and as long as the results from such tests are directly applicable to theories of organizational behavior, then theories of firm behavior are assured of a strong empirical foundation.

Chapter 15

TOWARDS A THEORY OF MARKET BEHAVIOR

Under classic conditions a market is analyzed in terms of demand and supply schedules and their intersection at equilibrium. While it is hard to imagine a market which does not consist of at least one buyer and one seller, it is evident that the notion of an equilibrium is extraneous to an analysis of decision behavior. Since the interactions that take place in a market are the result of the decision processes of both buyer and seller, it would seem reasonable to hypothesize that market behavior can be explained by an understanding of the separate decision processes and the ways in which they interact.

In the preceding chapter a set of procedures were described by which theories can be constructed to explain the decision-making processes of both consumers and firms. Such theories represent the decision-maker as having a set of decision processes which act upon and react to information which is already available to him in his memory or is made available by his environment. All behavior, under this theoretical framework, is a consequence of some describable decision process acting upon some ascertainable set of information. Concurrently, it has been argued that the decision behavior of individuals, as well as collections of individuals, and organizations or groups of individuals can be represented in a similar fashion. That is to say, whether one is dealing with one or many individuals acting by themselves or in groups, the resulting decision behavior can be described by a set of decision processes acting upon the relevant information. Since both individuals and firms frequently buy and sell

commodities through the medium of a market the behavior of the market must be a direct consequence of the individual decision processes.

Usually it is the variation in the price and the quantity supplied and purchased that constitutes what is known as a market's behavior. At any one period of time only one price^{1/} is in effect for each item. But over time these prices change, and it is this change in price that constitutes the market's behavior. At the same time, it is the change in price that must be explained if one is to be able to explain and predict the behavior of one or more markets.

In many markets the price per item is part of the information required by the individual or firm in order to decide on the quantity to buy or sell. As such the price is part of the decision-maker's initial conditions prior to making a decision. While the price may well change over time, the price at the moment is the item which is processed by that decision-maker. In these cases, the price is not subject to direct negotiation between buyer and seller. The buyer (seller) can decide to buy (sell) more or less of a particular item at the stated price but there is no opportunity to revise the price while this decision is being made.

A consumer in a department store, supermarket, or any other retail establishment is an exemplar of such activity. All items have a stated price

^{1/} Clearly the price for any one item can differ for wholesale and retail sales as well as for wholesale or retail purchases. But at any one period of time there is only one of each of these prices in effect in a specific market.

and the consumer's problem is to decide how much of each, if any, to purchase. In order to explain the consumer's behavior, all one needs to know are the prevailing prices and his decision processes. It is not necessary to know anything about the mechanism by which these particular prices are set. In some situations it may be necessary to know something about the recent history of the prices of these items, e.g., are they special sales prices? Even in this event, however, to explain the consumer's behavior it is quite unnecessary to know why the prices have changed.

On the seller's side of the market an example is provided by decision processes which account for the setting of prices in a department store. Again, at each moment of time there is only one price attached to each item in the store. And it is up to the price setter to decide whether to alter these prices or not. Such alternations, however, do not take place from instant to instant. They are based on a set of decision rules which are activated by certain events--notably, the recent history of sales, the level of inventories, the change in seasons, the approach of holidays, etc. All this information constitutes part of the initial conditions for the price setting decision process. Although, prices do change over time, the prices at any one period of time are explained solely by means of this process and not by a process which incorporates the customer's immediate reaction to these prices.^{2/}

In brief, under these conditions a classical market, with its own mechanisms for setting and adjusting prices, does not exist. Prices are set

^{2/} For a detailed model of the price setting decision process in a department store, which has survived empirical tests, see: R.M. Cyert and J.G. March, op. cit., Chapter 7.

by one set of decision processes and purchase decisions are determined by another set. At no one point in time do these processes directly interact. That is to say, the department store or supermarket is perhaps a convenient place for consumers to examine the available goods and for merchants to display their wares. But within these shops all purchases and sales are conducted at set prices and there is no opportunity for the classic balancing of prices and quantity to be carried out from one moment to the next. To understand the behavior of the buyer or seller, therefore, it is sufficient to know the decision processes by which each decides how much of each item to buy or what price per item to charge. Moreover, procedures have already been outlined by which theories of such behavior can be constructed and tested. Consequently, to account for this class of market behavior it is not necessary to develop a further set or body of theory. Manifestly, it is sufficient to be able to explain the behavior of the individual participants.

1. Price Behavior in a Security Market

There are other types of markets, however, in which buyer and seller come together and by their interaction directly establish a price and the quantity to be purchased. One such case is provided by the various security markets. In this instance the commodity in question, whether it be a bond, a stock, or a future, is known to both buyer and seller, and it is through their interaction that purchase and sales agreements are made. Since it is the fluctuation in the prices that is one of the chief characteristics of these markets, it is here if anywhere that a theory of market behavior is needed. Indeed, if it is the function of a theory of market behavior to explain, among

other items, the movement in prices, then the price fluctuations of the security markets are prime candidates for explanation by such a theory.

It is my position that in order to explain the behavior of security prices a theory of market behavior, as such, is not required. For even in this situation the behavior of the prices is a direct consequence of the decision processes of the individuals concerned, and no additional mechanism or theory is required to account for this behavior. Although classical theory employs a supply-equal-to-demand relation to establish an equilibrium market price, it is my assertion that not only is such a mechanism untestable and hence empirically vacuous, but it is also completely unnecessary. In brief, I am suggesting that the behavior of prices can be explained without reference to an equilibrating process. And further, that market behavior is strictly determined by the decision processes of the individual participants.

While this is hardly a novel conclusion, in that it is a somewhat obvious statement of the case, it implies for any specific market that one needs to know in detail the decision processes of all participants. If the behavior of certain commodity prices is being examined the number of such participants could be very large indeed. In addition if one has to be able to describe each of these decision processes the explanation of the behavior of the prices will indeed be a formidable and wearisome task. Security markets like other types of markets, are not composed of a collection of individuals indiscriminately competing for the opportunity to buy and sell. On the contrary, the process by which orders to buy and sell are executed is governed by certain institutional constraints, and the participants in the market can be classified into different categories. For example, actual transactions are usually conducted through

official agents, such as brokers and traders, and the participants can be categorized as to whether they represent investment societies, banks, insurance companies, pension funds, or private individuals.^{2/} Now, if the traders in a particular market behave according to a specific set of decision rules, then, and this is clearly a testable proposition, it is possible to describe the decision processes which determine their decision behavior. Similarly, if each category of investors behaves in recognizably different ways, such discrepancies must be a result of differences in their decision processes. Accordingly, if within each category decision behavior is sufficiently similar, then a set of decision rules can be described which will represent the decision-making procedures of each class of investors. Under these assumptions, all of which can be analyzed for their empirical validity, the problem of explaining price behavior becomes relatively simple and straightforward. For the prevailing price at any one moment will be a direct consequent of the trader's and the remaining, appropriate decision processes.

A. The Trader

In order to illustrate these remarks consider a recent investigation into the decision processes of the over-the-counter trader.^{3/} In the over-the-counter market--a market which accounts for approximately three-fourths

^{2/}While this is hardly an exhaustive set of categories, the participants in any market can be classified into observable sets of different types of investors.

^{3/}R.A. Jenkins, "Professional Trader Price Quoting in the Over-the-Counter Stock Market," unpublished Master's thesis, School of Management, Massachusetts Institute of Technology, 1964.

of the gross value of all security sales in the United States--the trader is responsible for quoting specific prices on all stocks in which he trades. Each trader maintains an interest in a particular set of securities, usually between 15 and 20 stocks, and in response to an inquiry will quote either a selling (asked) or buying (bid) price on any one of these securities.^{4/} Since the trader's price at a particular moment of time can be the market price--if a transaction is consummated, this price is the market price at this moment--one has to be able to explain the trader's price setting process if the behavior of prices over time is to be explained.

The trader is undoubtedly influenced by many different items of information. For instance a single trader has access to a number of sources of information, e.g., the Dow Jones ticker, the Dow Jones broad tape, the daily publication of the National Quotation Bureau which gives for each security the trader and the prices at the middle of the preceding day, as well as telephone conversations with other trades and stock brokers. Nonetheless, all trading activity is carried on over a telephone in very brief intervals of time. Accordingly, at any one moment a trader can be asked over the telephone for the price on a particular security. He responds, as a rule, with the bid and asked prices on a hundred-share lot. If this price is accepted, a transaction has been made and the trader has either sold or bought a number of hundred-share lots.

It should be noted that the trader only deals with stock brokers or other traders. Under no circumstances is it possible for a private individual

^{4/} The difference between the asked and bid prices on a security is what is known as the spread.

or institution to deal with a trader directly. The stock broker takes orders from private or institutional investors and then telephones a trader to ascertain price. Since the broker charges a fee for this service, the price to the ultimate purchaser differs somewhat from the price set by the trader. Moreover, it follows from this outline of the procedure that when a trader receives a telephone call he knows that the broker has an order to buy or sell. Thus, whether there will be an immediate transaction or not depends entirely upon the broker's reaction to the trader's quoted price. Since the broker can telephone any of the traders who are known to have an interest in this particular security, he is not dependent upon a single quote from one trader. However, as soon as the broker accepts a price that is the price at which the transaction is made, and consequently it is the market price in the particular security at that instant of time.

Before examining the trader's pricing decision process in detail it is pertinent to consider his possible alternative strategies. One alternative is for the trader to deliberately maintain either a net long or net short position in a particular security. In a rising market the value of his inventory will increase, and as a result he would want to have a net long position. Conversely, in a falling market a profit can be made by buying back stock at a lower value than that which he sold it for. Accordingly, he would want to maintain a net short position. While during certain periods of time traders may actively seek to maintain long or short positions^{5/} the

^{5/} The most notable period when these strategies were actively pursued was in the latter part of the 1920's, see: I. Friend, et. al, The Over-the-Counter Securities Markets, McGraw-Hill, New York, 1958.

current strategy is to make a profit by trading on the difference between the bid and asked prices. Although traders may make a certain amount of profit by taking advantage of a position they find themselves in, the principal monetary return comes from buying at his bid price and selling at the asked price. As a result, in order to be successful the trader must maintain this spread between prices such that when combined with the volume of trading an adequate level of compensation is assured.

B. The Pricing Decision

Given this brief description of the trader's function in the over-the-counter market, it is now relevant to examine the pricing or quoting decision process itself. A decision is required of a trader each time a broker telephones to ask for a price. Since the trader must reply virtually immediately, one would not expect the pricing process to be unduly complex. According to the study mentioned above^{6/} the basic components of the pricing process can be represented as follows:

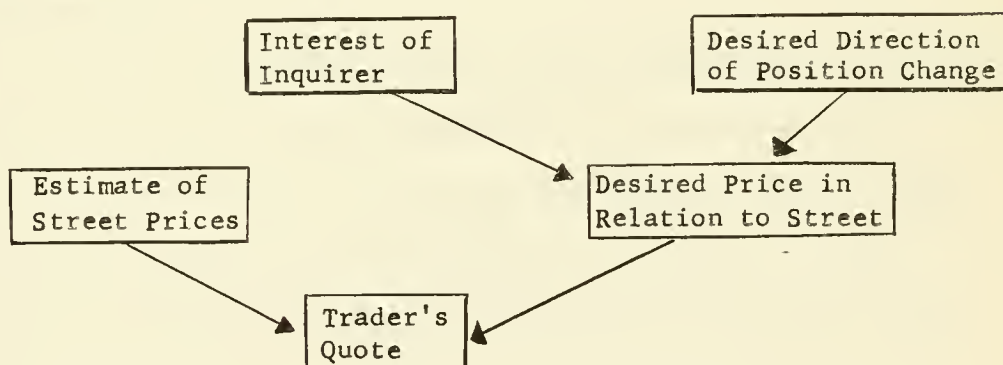


Figure 1

^{6/} R.A. Jenkins, op. cit.

While each of these components is a result of the influence of a number of other factors,^{7/} the final decision process which takes place at the end of a telephone can be represented by the interaction of these four items. For example, a trader alters his quote depending upon the characteristics of the inquirer. Such factors as whether the inquirer is a buyer or seller, whether the orders from this person are usually large or small and whether he is a friendly^{8/} competitor or not affect the quote in a manner to be outlined below.

At the same time the trader knows whether he wants to increase or decrease his current long or short position in a particular stock. For at all times the trader knows his current position as well as his estimate of the position he would like to have. Since traders normally have a maximum amount of money that they can invest in any one security, their general impressions and attitudes toward the market, constrained by this limit, are what identify the position he would currently like to be in. Any discrepancy between the desired and actual position provides what has been labelled the desired direction of position change.

The estimate of the street or current market price is derived by the simple process of listening to the reply on the telephone. If the trader's quote is accepted then he is either right on or a little low (on asked price), right on or a little above (on bid price) the current market. Conversely, if no transaction is effected, his asked price is a bit high and his bid price is a

^{7/}For a full description of the decision process see ibid, Chapter 3.

^{8/}A friendly competitor is one who does not take advantage of a bargain or poor quote.

bit low. If, for some reason, the stock has not been traded for awhile, an individual trader can obtain an estimate of the current price by telephoning a competitor. But if the stock is being actively traded, each trader will have a fairly accurate estimate of the current market price. Given this estimate and any desired change in position, the quoted price can be directly determined.

While the actual increments, e.g., $1/8$, $1/4$, $1/8$, etc., may vary with different securities^{9/} the price setting decision process can be represented by the following table:

Inquirer and His Interest	Desired Direction of Position Change	Desired Price Relation To Street	
		Bid Price	Asked Price
Retail Buyer	longer	$1/8$ above street	$1/4$ to $1/2$ above street
	indifferent	equal street	$1/8$ above street
	shorter	$1/4$ below street	equal street
Retail Seller	longer	equal street	$1/4$ above street
	indifferent	$1/8$ below street	equal street
	shorter	$1/4$ to $1/2$ below street	$1/8$ below street
Friendly Competitor (Interest unknown)	longer	equal street	$1/8$ above street
	indifferent	equal street	equal street
	shorter	$1/8$ below street	equal street
Enemy Competitor (Interest unknown)	longer	much lower	much higher
	indifferent	than street	than street
	shorter		

Table 1

^{9/}See ibid. Chapters 3 and 4 for a detailed discussion of the variation in spread.

This table describes the components of the price quoting decision process in sufficient detail to permit some of the processes to be subjected to test. Further, from the evidence presented in the study, these decision processes are sufficient to account for a substantial proportion of the observed changes in traders' prices for a number of securities.^{10/} Consequently, it can be accepted, for the moment, as a detailed representation of the price setting decision process.

Of particular interest in this decision procedure is the mechanism by which a price is changed. If the trader quotes a price which does not result in a transaction, no change is made in the price. But, if a transaction is effected--i.e., the broker accepts the trader's price--then the trader's price will change in the direction specified by the process outlined above. As a result, price changes are for the most part a consequence of a transaction being consummated and are seldom altered to effect a transaction. Thus, prices respond to the occurrence of transactions--and are in effect determined by these contracts.

Lest the reader feel that somehow the price setting process could not be as simple as portrayed above, or that it would be more likely for the trader to change his price in order to get transactions, it is worth noting that the process outlined above apparently reflects a decision procedure which is used by many people when placed in roughly the same situation. That is to say, when faced with the task of bidding for contracts in an experimental market

^{10/} See ibid. Chapters 4 and 5.

most subjects employ decision procedures which are strikingly similar to those used by the over-the-counter trader. This observation is one result of the experimental study of problem solving control mentioned earlier in this book.^{11/} The experiment itself consists of placing a subject in a situation where he has to announce bids in two markets simultaneously. The subject states his bids in monetary terms, and the experimenter by consulting a specific list of random numbers determines whether these bids "win" or "lose." A bid "wins" when it is below the experimenter's number, and "loses" when it is equal to or above. There is a fixed cost associated with each trial and the subject is restricted to making at most one new bid on each trial. Hence, on each trial the subject has to decide which market to leave alone and which bid to alter if at all. A subject's earnings are a direct function of the contracts he wins over a given number of trials.

In this experiment a subject's behavior is a record of bids or prices on two markets. These prices change over time. Hence, an explanation of this behavior consists of an explanation of the changes in the respective prices. Since subjects have no direct knowledge about the list of numbers employed by the experimenter their behavior is clearly a function of how they decide to respond to their record of wins and losses as it unfolds. While many of the subjects who participated in this experiment employed slightly different decision procedures, there is one principal set of processes that characterizes and accounts for a large proportion of the observed behavior. This process is expressed by the following table:

^{11/} See W.F. Pounds, op. cit., Chapters 4 and 5.

	WIN	LOSE
WIN	Raise the lower of the two bids	Lower the losing bid
LOSE	Lower the losing bid	Lower the higher of the two bids.

Table 2

Although this set of processes does not include the amount to alter the price by, nor a procedure for deciding what to do in the event both markets have won or lost and both bids are the same, it does contain the principal components of most subjects' price setting procedures. The principal characteristic of the process, aside from its simplicity and symmetry, is that new bids are made in response to contracts made or lost. Prices are lowered when losses occur and are raised or held the same when contracts are won.

The significant point is, of course, that it would be perfectly simple to choose prices according to some sampling or other statistical procedure. To employ a notion of sampling one would note the frequency of wins and losses at various prices in each of the markets and choose that price which appeared to yield the desired earnings. In fact, despite the statistical training of many of the subjects, very few chose to behave in this or any other fashion. As a result, it appears that when little or nothing is directly known about the behavior of the environment, processes are frequently employed which respond to rather than anticipate the occurrences of the relevant events. This is not to suggest that the subjects in this experiment are all fledgeling traders, nor is it being suggested that the two situations are the same. But the similarity in characteristics of the traders' and subjects' decision

processes is too striking to ignore. And since the simplicity of the traders' pricing decision process is reflected in the bidding process of the subjects, the empirical validity of the trader's price setting process has received a certain measure of independent empirical support.

C. The Broker

In the over-the-counter market the function of the broker is to accept orders from customers and by talking directly with the traders negotiate the transactions. Clearly, the broker does not have to accept the first price he receives over the telephone. But if he frequently deals with a particular set of traders he in turn will have an estimate of the relation between their prices and the prices of other traders, i.e. the street. What the broker does not know is the trader's desired direction of position change, and hence whether his price is deliberately slightly above or below the street price. The broker's task is to find a favorable price for his customer, and if he believes he can do better by trying another trader all he has to do is pick up the telephone and find out.

One of the factors which influences the trader's price, not noted above, is the activity or volume of purchases or sales in a particular security. Each trader has a ceiling on the amount of money he can commit to a single stock, which given the prevailing price places a limit on the number of shares of this stock that he can hold. Now, if the traders in Stock A are known to be holding approximately 500 shares each, and the broker receives an order to buy 4,000 shares, he is clearly placed in a bit of a dilemma. Since no one trader can fill his order, he must buy (or sell) from a number of traders. News of this activity in Stock A will spread to competing traders fairly

tests corroborate the theory, then this particular set of decision processes would represent in detail the procedures by which investors of trust funds select securities for their portfolios. Once these procedures are known the only other items of information required are the amount of funds available for investment classified by the types of portfolios desired, e.g. growth, income, etc. By an application of the decision process to current market data specific portfolios of securities are generated. These portfolios represent the orders which are given to the broker by the investor. As a result, it is these portfolio decisions which constitute the origin of the broker's orders.

It is worth noting that portfolio decisions are relatively insensitive to the exact prices prevailing in the market at the time the portfolios are selected. Since the actual price for a particular order is only determined after the broker has received it and has contracted with a trader, the investor must select his portfolios on the basis of some previous prices. While these prices may closely approximate the actual prices paid after the broker has completed his transaction, nevertheless, portfolio decisions are clearly made without an exact knowledge of the price per security that will be paid.

Due to various legal constraints investors of trust funds are not allowed to purchase securities on the over-the-counter market. Hence, with respect to this market a knowledge of the trust investment process does not provide the basis for one category of investors. However, since it is possible to describe the portfolio procedures of trust investors there is no reason to suppose that the investment behavior of other institutional investors who do participate in the over-the-counter market cannot be described in a similar manner. Further, the theory of decision behavior outlined in previous chapters

provides the theoretical schema by which these investment processes can be described and the particular selections explained. Consequently, since a theory of each class of investors can be constructed and tested, it is clearly possible to describe the processes by which the orders received by brokers are generated.

2. Testing the Market Processes

In the case under consideration the market for securities consists of the interactions of brokers and traders. If the traders' decision process is accepted, for the moment, as it is given above, and if a simple decision process was constructed to account for broker behavior, the behavior of prices would be determined by these two processes. That is to say, if one is not concerned about explaining the flow of buy and sell orders to the broker, all that is required is the sequence of orders plus the two decision procedures. With the orders as part of the initial conditions, the behavior of the relevant prices will be a result of the interaction of the broker's order contracting process and the trader's price setting process.

Manifestly, it is possible to examine the simple case where there is only one trader who holds an inventory in a particular stock. Since this condition is likely to occur only when there is little interest and activity in a security, the number of brokers who receive orders for this stock will be quite limited. Hence, the behavior of the price of this security will be a direct consequence of a few brokers interacting with one trader. Given such a situation, it is neither difficult nor laborious to determine the particular decision processes employed by each of the participants. Once these processes

are described, with the brokers' orders forming a part of the initial conditions, the behavior of the price of this particular security can be immediately explained. For the interaction of these decision processes, if they are each able to account for their respective decision behavior, will generate a sequence of price movements which should be identical to the observed.

In order to test the accuracy with which this model of market behavior reproduces the observed movements in price, it is only necessary to set up a criterion of success and failure and compare the two time series. Such a comparison can be conducted upon the actual prices themselves, as well as on whether the model produces a set of prices that move at each decision point in the same direction as the observed. Once measures of success and failure are defined--i.e., under what conditions the model's price is to be considered the same as the actual--the model's level of success can be measured by the frequency with which it accounts for the observed price change. Since each of the individual decision processes can be independently subjected to test, the model as a whole can be satisfactorily tested on its ability to reproduce the observed time series by determining its relative frequency of success.

In a situation where there is more than one trader who holds an inventory in a particular security the model would become correspondingly more complex. For once there are several traders as well as a number of brokers there may be more than one price prevailing at any one point in time. Each broker agrees to a transaction when he thinks he has secured a favorable price. But each broker does not canvass all traders before making a decision. In addition, more than one broker may be interested in a certain security at one period of

time. Therefore, it is possible for there to be slightly different prices prevailing at one instant of time.

In order to reproduce these detailed events the model would have to include the individual decision processes of each participant. To empirically determine these separate processes would be a time consuming task. But if a complete explanation of a certain stream of price behavior is desired the separate decision processes must be taken into account.

However, if an explanation of each movement in price is not required and if the behavior under investigation is concerned only with some of the more aggregate characteristics of the price changes over an interval of time, e.g., direction of change from beginning to end of interval, incremental change, etc., then a simplified model would suffice. Such a model could perhaps consist of a generalized broker's decision process interacting with a generalized price quoting process. Whether such a model would produce the desired behavior is open to empirical investigation. But since each of the individual processes can also be independently subjected to empirical test, the empirical validity of the entire model is not solely dependent upon the general characteristics of the generated time series being similar to the observed. Consequently, it would appear that it is quite possible to develop a general model of price behavior without too much difficulty.

The point to note is that none of these models require an equilibrating mechanism. Each is based solely upon the interaction of independent decision processes. Thus, although their empirical validity has yet to be demonstrated, the research described above is in my opinion sufficient to indicate the theoretical and empirical merit of this approach. Accordingly, while only one

type of market has been examined in any detail it appears that all market behavior can be explained by theories which include the decision processes of the individual participants and which do not incorporate the classic equilibrating hypothesis.

Chapter 16

BEHAVIORAL THEORY, MICROANALYSIS AND POLICY DECISIONS

In the latter part of Chapter 10 it is noted that the only condition under which it is possible to test econometric hypotheses is when there exists an independent method of measuring the population's stability. Within the confines of econometrics itself it is not possible to determine whether each sample comes from the same underlying population--i.e. whether the population's characteristics remain stable over the period in which the samples are drawn. Accordingly, unless the behavior of the population can be separately determined and the constancy of its relevant characteristics ascertained, econometric hypotheses cannot be submitted to a process of refutation by empirical test.

The previous four chapters, however, have been concerned with describing and examining a theory as well as an experimental method by which the behavior of an individual, a class of individuals, and a group or organization of individuals can be explained. While a general theory of economic decision processes has yet to be fully developed, the basic theory of decision-making behavior provides the structure around which such a theory can be constructed. Further, the various models of this theory which have already been successfully submitted to test demonstrate the empirical testability of these behavioral theories. That is to say, the evidence strongly supports the proposition that testable theories can indeed be constructed which are able to describe and explain the economic decision behavior of individuals whether acting singly or in groups.

If the proposition is accepted the question then arises as to whether these behavioral theories can serve as the independent measure of a specific population's characteristics and stability. For example, if a theory is

developed which explains a particular sequence of economic decision-making behavior, then this theory can be employed to determine the point at which this decision behavior begins to change. Such a point occurs the minute the theory is no longer able to describe and explain the observed behavior. In other words, if a theory of a certain decision process has survived a number of tests and has already been shown to be sufficient to account for this behavior, then as soon as the theory can no longer account for the behavior, the inference can be drawn that the economic decision process has altered in some way. Accordingly, up until the moment that the theory can no longer account for the observed behavior, the theory itself provides a means for determining both the constancy and the characteristics of the decision process.

Behavioral theories represent economic decision processes. Since their hypotheses can be subjected to test, the constancy of specific decision process is assured by the ability to detect change. Once a theory fails to explain, a change can be presumed to have occurred and the constancy of the decision processes is placed in doubt. But until this moment arrives the theory itself describes the relevant decision process, and is a direct and independent measure of their stability over time.

Once a theory fails to explain the occurrence of a specific event, its hypotheses are re-examined in order to detect the source of the error. If after further tests and analysis the theory is once again sufficient to account for the relevant decision behavior, then the alterations in the theory represent the changes that are presumed to have taken place in the observed decision behavior. Accordingly, not only can a behavioral theory represent the actual decision processes at one period of time, but as it is amended it can also account

for new observations. Consequently, the amendments are a record of the changes that are occurring in the relevant decision behavior.

Since a behavioral theory describes observed decision behavior in terms of decision processes and their attributes or component parts, behavioral theories also provide a description of the characteristics of decision processes. As a result, any alterations to a theory must be reflected in the hypothesized processes and their attributes which in turn alter the characteristics of the decision-making process. While some changes may verge on the insignificant, others may considerably affect the resulting decision behavior. In either event, however, the existence of a testable theory of this decision behavior provides both an indicator of change as well as a method whereby the extent of the change can be assessed. Consequently, it appears that behavioral theories can indeed serve as an independent measure of the constancy as well as the alterations in specific economic decision processes.

For example, consider the decision behavior of an investor of trust funds. It is clear from foregoing chapters that a theory can be developed which describes such an economic decision process in considerable detail. Further, such theories can be submitted to test and on the basis of such tests can be classified as being sufficient to account for a particular sequence of portfolio selections. As long as the theory continues to survive empirical tests, its hypotheses consist of a precise specification of the factors that are considered as well as the order in which they are related in the selection of securities for particular portfolios. Accordingly, both the discrimination nets and the remaining processes represent the attributes and characteristics of the portfolio selection process. If a change occurs such that some of these hypotheses--notably, the

respective discrimination nets--need to be amended, then the extent of the change in the observed behavior can be measured by the number and type of alterations made in the theory's hypotheses. In addition, the new theory now serves as the basis from which further changes in the observed behavior can be detected, analyzed and assessed. Therefore, a theory of the decision processes of trust investors provides the requisite hypotheses and observable initial conditions to establish explanations of the behavior of this class of economic actors.

1. Behavioral Theories and Econometrics

Since behavioral theories are a direct representation of economic decision processes the next point to examine is the relation between econometric and behavioral hypotheses. For if behavioral theories are to be employed an independent measure of economic decision behavior these must be a direct connection between these two types of theories if this measure is to perform the required task. That is to say, unless econometric hypotheses refer to the same attributes and characteristics of the observed behavior as the corresponding behavioral theories, the existence of testable theories of economic decision processes will not enable econometric hypotheses to be submitted to test.

Traditionally econometric hypotheses are formulated by reference to the theories and hypotheses of classical economics. Accordingly, econometric hypotheses usually contain concepts and terms which correspond directly with the terms and concepts employed by classic economics.^{1/} Since the theories from

^{1/}For specific examples and further detail see the econometric models described in Part III.

which these concepts are taken cannot be subjected to test it is highly questionable whether many of these terms are themselves subject to empirical analysis.^{2/} In addition, it is apparent from the analysis in Part III that even if some or all of the terms in an econometric theory refer directly to observables the empirical validity of the theory cannot be assessed due to the lack of appropriate measures on the underlying population. Thus, although the objective is to transform econometric hypotheses in to a state of empirical testability, it appears that in order to accomplish this task the method by which such hypotheses are formulated must also undergo a change.

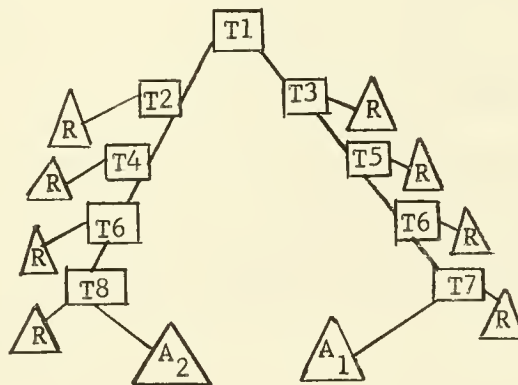
One possible method of approach is to construct econometric hypotheses on the basis of the corresponding behavioral theory. Behavioral hypotheses are deterministic statements which represent the components of the relevant economic decision-making process. Econometric hypotheses, however, are stochastic propositions which usually consist of a linear relation among the appropriate variables. To formulate an econometric theory on the basis of a behavioral clearly requires, among other things, that determinate relations be converted into a statistical framework. If the resulting econometric theory were to be used to explain the behavior of a single economic actor, there would be little reason to justify such a transposition. For if the observed behavior can be accounted for by a deterministic theory the dictates of parsimony would preclude the addition of unnecessary statistical factors. But, if the

^{2/}For a detailed examination of the empirical content of the concepts employed in the classical theories of utility and demand see: G.P.E. Clarkson, Theory of Demand, op. cit., Chapter 4.

econometric theory were to be employed to account for the aggregate behavior of a class of economic actors, the transposition to statistical hypotheses may indeed simplify the resulting theory.

In order to construct such econometric hypotheses it would be necessary to begin with a tested theory of the decision behavior under consideration. One of the principal features of decision-making theories is the discrimination nets with which the relevant information from the memory and the environment is processed. These discrimination nets are composed of a series of tests or operations. Since these tests or operations are represented in the nets by their names it is clearly possible to identify one variable with each of these items. A particular discrimination net would then be represented in econometric terms by a list of variables. If the discrimination net in question were a simple sequence of tests, it could then be represented by a linear relation of the respective variables. Consequently, a statistical hypothesis could be constructed by forming a linear relation of the variables in the discrimination net with the addition of an error term.

For example, consider the following discrimination net which is composed of eight tests or operations.



If each item is represented by a different variable the discrimination net could be transposed into the following econometric hypotheses:

$$y_1 = \alpha_1 x_1 + \alpha_3 x_3 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_7 + u_1$$

$$y_2 = \beta_2 x_2 + \beta_4 x_4 + \beta_6 x_6 + \beta_8 x_8 + u_2$$

For in this case the cells marked $\underline{A}_1$ and $\underline{A}_2$ are not reached unless each of the respective tests is passed successfully.

Suppose for the moment that this discrimination net represents a particular segment of a specific economic decision process. Suppose further that by the addition of the requisite information (initial conditions and interpretive rules) this net is corroborated by empirical test--i.e., it is sufficient to account for the observed behavior of a number of individuals. As such the net is a general theory of this particular sequence of decision-making behavior. The net is tested by applying it to the observed behavior of single behavior of single individuals. Thus, even though it is possible to employ this net to explain the behavior of a number of individual decision-makers the net itself is not constructed to be tested against aggregate measures of such behavior. Manifestly, it is to account for these aggregate measures that the econometric hypotheses are to be employed. That is to say, by transposing the discrimination net into the corresponding econometric relations hypotheses are established which are suitably constructed for testing against aggregate data. For data can now be collected on the decision behavior of this class of economic actors and the coefficients can be estimated by the customary statistical procedures.

Once these relations are established with the appropriate values for their coefficients and error terms they can then be employed to explain or predict other collections of similar aggregate data. For, by constructing these statistical relations from the tested hypotheses of behavioral theories a method is provided for measuring the stability and characteristics of the underlying population. As long as the behavioral theory is continually subjected to and survives empirical tests, the behavior of the total population can be assumed to remain unchanged. Consequently, all samples drawn during this period come from the same population, and the basic requirement of statistical testing is satisfied. If during a sequence of tests the behavioral theory fails and has to be amended in order to continue to account for the observed behavior, then such changes as are made must be reflected in the corresponding econometric relations. Once some tests are deleted while others are added to form new discrimination nets, the same alterations must be made to the respective variables. For, it is only as long as the two sets of theories remain structurally similar that the empirical testability of the econometric relations can be assured.

In order to illustrate this method of procedure in more detail consider the microanalytic model of the household sector referred to toward the end of Chapter 10. As noted above^{3/} this model is concerned with the demographic and economic behavior of household units. In particular stochastic relations are derived which represent what are called the operating characteristics of

^{3/}See Chapter 10, section 4.

the household. These relations are developed from sample survey data and refer to the probabilities of the behavior of some 23 dependent variables. The behavior under investigation is that of household spending on certain durable goods.^{4/} Although the data are based on 3,000 units for each year, this microanalytic model is confronted with the same empirical obstacles as any other set of econometric relations. For unless it is possible to independently determine the stability and characteristics of such household behavior, it is not possible to subject these relations to a process of disconfirmation by empirical test.

To transpose this microanalytic model into an empirically testable state, according to the proposed procedure given above, it is first necessary to develop a behavioral theory of this household behavior. Since the microanalytic model is concerned with purchases of certain durable goods, it follows that a behavioral theory must be constructed which can account for the decision-making process of households with respect to home ownership, monthly rent, car ownership, purchases of household durables, etc. In brief, a theory is required which describes and explains this segment of consumer behavior.

In Chapter 14 an outline of such a theory is described. It will be recalled that the proposed theory of consumer behavior contained a number of basic sets of decision processes. The first of these is a decision procedure for allocating the consumer's (or in this case the household's) total income among the various

^{4/}

A list of the variables appears on page 215.

categories of commodities, such as housing, food, clothes, entertainment, etc. The theory assumed that the proportion of total income allocated to each of these categories would remain fairly stable over time, and would also correspond to the purchasing behavior of other consumers in the same social and economic class. While evidence was not presented to support these propositions it is clear that they can readily be submitted to empirical test.

The next set of decision processes are concerned with the procedures by which the funds assigned to each commodity category are allocated over the actual items purchased. In particular these processes were postulated to include: a decision procedure which determines whether purchases are to be made by cash or cash equivalents or by a series of periodic payments; decision rules which describe the processes by which the consumer (household) selects one set of commodities, within a particular category, from the available alternatives; decision processes that permit the theory to adjust its selection behavior in accordance with certain expectations about the future behavior of prices and other variables that are considered important; and a procedure for resolving such conflicts as might arise among the primary allocation of funds to the respective categories.

Manifestly, a theory can be constructed which accounts for the decision-making processes of households. If one is solely concerned with expenditures on durable goods, the theory can be restricted to such commodity categories. The point to note, however, is that once such a theory has been developed and tested its decision processes will provide the basis for the desired microanalytic model. For, while such a behavioral theory may appear to the reader as somewhat complex, it should not be forgotten that its decision processes

are represented by discrimination nets. And, as is argued above, the components of discrimination nets can be classified as variables which can then be composed into the appropriate econometric relations. Consequently, once a testable theory of household decision behavior with respect to durable purchases is constructed a microanalytic model can be developed which is similar in attributes and characteristics to the behavioral theory. As long as the behavioral theory continues to account for observed behavior the resulting microanalytic model can be employed to explain and predict. For, once again a knowledge of the population's decision processes provides a direct measure of the stability and characteristics of these decision procedures. Also, the similarity in structure between the behavioral and microanalytic relations permits the behavioral theory to serve as a direct measure of the microanalytic model's population--a measure which is vitally required if statistical tests are to have any empirical significance.

Since the microanalytic approach is designed to be able to employ the data from sample surveys and other interview techniques, there does not appear to be any reason why such data cannot still be employed to estimate the relations of the revised models. The only difference would be that instead of basing the econometric relations upon the survey data--i.e. constructing relations from the data reported in these surveys--the microanalytic models would now be developed from the behavioral theory. As a result, the survey data would be used to estimate the already specified relations. Once this is accomplished further data can now be explained or predicted in the normal fashion.

One possible consequence of this approach to microanalytic model construction is that these revised models may well contain variables for which

sample data has not yet been collected. While the occurrence of such an event would delay the process of submitting the model to empirical test it could hardly be classified as a serious obstacle to the whole endeavor. The remedy, of course, is to commence a survey which will yield the appropriate data. Moreover, as there is little to be gained by the indiscriminate gathering of data such a theoretical focus would provide a much needed structure to the data collection process.

2. Time Series and Behavioral Analysis

Another way in which behavioral theories can be employed to further econometric theory is in the statistical analysis of the behavior of time. One class of time series that have received a considerable amount of attention is the set which reflects the movement of security prices over time. The object of many investigations has been to determine whether the movement in prices follows some detectable pattern or whether such price behavior is indistinguishable from that which characterizes Brownian motion or a random walk.^{5/} Recently, results have been published which suggest that although the behavior of stock prices does not appear to be consistent with a pure random walk, stock price behavior is consistent with the hypothesis that prices

^{5/}Some recent examples of this endeavor are to be found in: M.F.M. Osborne, "Brownian Motion in the Stock Market," Operations Research, Vol. 7, March-April, 1959, pp. 145-173; H. Working, "Note on the Correlation of First Differences of Averages in a Random Chain," Econometrica, Vol. 28, October, 1960, pp. 916-918; and H. Houthakker, "Systematic and Random Elements in Short Term Price Movements," American Economic Review, Vol. 51, May, 1961, pp. 164-172. (The Osborne and Working papers are reprinted in: P.H. Cootner, ed., The Random Character of Stock Market Prices, M.I.T. Press, Cambridge, 1964, pp. 100-128, and 129-121, respectively.)

move in a random fashion between certain limits or reflecting barriers. While these barriers do not remain constant over time, the shifts are what indicate a trend, the movement within the barriers is indistinguishable from that of a random walk.^{6/} Of course, if stock prices follow some discernible pattern then once this behavior is known to at least one individual his monetary reward would no doubt justly compensate the effort involved in reaching this discovery. However, it is presumed that the behavior of stock prices is not being examined solely for pecuniary motives, and that the fundamental objective of these investigations is to be able to explain the behavior of such time series by means of the appropriate statistical analysis.

The change in prices over time is a consequence of the interaction between broker and trader--i.e. by studying such time series one is investigating the final result of a set of market processes. While statistical investigations may well lead to stimulating results, it would appear to me that a good deal more could be learned about the relevance of certain characteristics of these time series by examining the processes by which they are generated. For, by an analysis of the price setting process and the sequence in which orders are placed it may well be possible to identify the principal factors which affect the movement of prices over time.

In the previous chapter a theory of the decision-making process of the over-the-counter security market is discussed. In the simplest case the time series

^{6/} See the results and analysis cited in: S. Alexander, "Price Movements in Speculative Markets: Trends or Random Walks," Industrial Management Review, Vol. 2, May, 1961, pp. 7-26; and P.H. Cootner, "Stock Prices: Random vs Systematic Changes," Industrial Management Review, Vol. 3, May, 1962, pp. 24-45. (Reprinted in: P.H. Cootner, op. cit., pp. 199-218, and 231-252, respectively.)

of the prices for one security is a direct result of the interaction between the trader's price-setting process and the broker's order-placing process. Since both of these processes can be described in detail, it is clearly possible to construct a specific model of this market behavior. If the model is provided with the requisite information the final result of its behavior will be a series of prices. This sequence of prices will, if the model has been properly constructed, correspond to an actual, observable time series--namely, the actual market prices for this security during the interval in question. Since the generated sequence of prices is a perfectly respectable time series it can be submitted to the same statistical tests as are employed in the above mentioned investigations of stock prices. But, and this is what appears to be the important point, whatever the outcome or inferences that are drawn from such statistical analyses, the model provides the mechanisms by which these statistical characteristics are produced. Indeed, if it can be shown that certain decision processes, on the part of both broker and trader, lead invariably to the generation of time series with particular statistical characteristics, then a basis would have been established from which decision processes might be inferred from statistical characteristics. If, as may well be the case, different sets of decision processes lead to time series with significantly different statistical characteristics, then these characteristics with their corresponding decision processes can be grouped into separate classes.

While these suppositions may appear somewhat idealistic, and may well be rejected out of hand as requiring far too much effort to investigate, permit me to remind the reader that they are perfectly straightforward propositions which can readily be subjected to empirical analysis. Further, whether they turn out

to be corroborated or not the process of subjecting them to test would not be as involved as it may at first seem. For once a theory of the market process is developed its behavior can be assessed under a variety of initial conditions and environments. If the time series so produced vary significantly in their characteristics, then it would appear that these characteristics can be classified by decision process as well as environmental conditions. Concurrently, by examining the decision processes governing the setting of prices in several markets it would also be possible, given some differences in these processes, to identify the effect these differences have on the behavior of prices over time. In short, such an investigation would lead to a more detailed understanding of market behavior as well as an explication of the origins of specific time series and their statistical characteristics.

Even though this book is primarily concerned with the development of an empirical science of economics, consider for a moment one application to which such knowledge of market processes could be put. Suppose that you are asked by some organization such as the Security and Exchanges Commission to give your considered opinion on the way in which the trading in securities on the major exchanges should be managed. Their concern is not so much with the administration and policing of their regulations. Rather they are concerned with the behavior of security prices and request you to consider the problem of how to secure an orderly market in such prices.

One way of describing the properties of an orderly market is in terms of the characteristics of the time series of its prices. If certain classes of time series are considered satisfactory by the commissioners the answer to their problem lies in prescribing the set of decision processes which will produce such a time series. While further investigations of market processes would undoubtedly

lead to amendments and refinements in the recommendations, the process by which such knowledge is acquired would remain the same as outlined above. Consequently, even though some might still be tempted to say that they had found the "best" way, the ability to empirically evaluate alternative proposals would enable "better" procedures to be discovered and adopted.

3. Behavioral Theory and Policy Decisions^{7/}

Historically, one of the principal concerns of economists has been, and will no doubt remain so in the future, to employ economic theory with its concomittant analytical tests to develop and prescribe policies for the many important public and private economic decision problems. Whether the problems concern the national welfare or pertain to private and individual enterprise, economists have always responded to policy problems with recommendations as to the most appropriate policies to be adopted in each of the problem areas. These policies are based either upon classical or econometric analysis. Accordingly, if the analysis contained within the second and third parts of this book is correct, it follows that such recommendations have been based upon empirically untestable theories. Indeed, while specific policy decisions may or may not have proved useful at the time, none of these decisions were based upon an empirically tested theory of the economic decision process under consideration.

^{7/} This section is indebted to: G.P.E. Clarkson, "Interactions of Economic Theory and Operations Research," in A.R. Oxenfeldt, (ed), Models of Markets, Columbia University Press, New York, 1963, pp. 339-361.

The existence of a method by which testable theories of economic processes can be constructed profoundly alters the task of formulating policy decisions. For, once a theory is established which describes and explains the economic behavior in question this theory is the basis upon which policy decisions ought to be made. Empirical knowledge of the relevant decision processes leads immediately to a knowledge of the principal factors which affect the decision behavior. Experimental investigations of these processes will provide a knowledge of the changes in the decision behavior which correspond to certain selected alterations in the environment or within the decision processes themselves. Manifestly, such a body of knowledge constitutes a sound empirical base upon which to construct policy decisions.

To illustrate the effect testable theories may have upon well known policy decisions consider the traditional conception of how industrial pricing policies ought to be regulated. Classical theory, as noted in Part II, asserts that competitive pricing is the most efficient method of keeping the prices of finished products, e.g., consumer prices, as low as possible. Accordingly, when competitive pricing appears to have vanished and one or two companies are observed to dominate an industry, antitrust measures are invoked with the avowed intent of restoring competitive pricing to that particular market. But if, as investigations of business behavior suggest,^{8/} the pricing decision is only one of a firm's decision problems, then to increase

^{8/}For further discussion of this point see: R.M. Cyert and J.G. March, op. cit., Chapter 6; and H.A. Simon, "New Developments in the Theory of the Firm," American Economic Review, Vol. 52, May, 1962, pp. 1-15.

the number of firms in the market may not have the desired effect. For unless it can be shown that the number of firms in the industry is an integral part of the pricing decision process, it does not make much sense to invoke anti-trust measures whose purpose is to increase the number of competing firms. For example, it has been observed in a number of cases that increases in internal administrative costs frequently lead large firms to centralize their decision-making processes.^{9/} In effect, these firms are led to replace internal pricing mechanisms with central planning. Departments no longer maintain their own profit and loss figures, but work instead from allocated budgets and set prices. A further stimulant to centralized decision-making is provided by the high-speed computer and management information and control systems; and it is clear that many firms are making use of this data-processing capability. If prices and budgets of large corporations are set by a central plan, this plan will not be sensitive to changes in the external environment. For the vast amount of coordination required by centralized decision-making precludes the possibility of these decision processes being too sensitive to external disturbances. Consequently, once prices within an industry are judged by some regulatory body to be too high and antitrust measures are invoked, then for these measures to be effective they must somehow directly affect some of the principal components of the centralized pricing process. Unfortunately, not enough is yet known about planning and pricing decision

^{9/}For example see: J.G. March and H.A. Simon, Organizations, Wiley, New York, 1958, Chapter 7; and H.A. Simon, The New Science of Management Decision, Harper and Brothers, New York, 1960, Chapter 5.

processes to suggest effective procedures for inducing the desired change. But the evidence is sufficient to call into question many of the traditional beliefs about the efficacy of antitrust measures in controlling prices.

As a further example of how policy conclusions may have to be revised consider the traditional conception of a firm's reaction to various tax policies. In particular what is the effect of levying a lump-sum or poll tax on a corporation's prices? As noted in Chapter 3 the assessment of a poll tax is supposed to be one of the most effective ways of imposing a tax upon a corporation without having the cost of this tax passed on to the consumer. The conclusion is based upon the classical assertion (derived as equation on page) that changes in fixed costs are to be ignored when making pricing and output decisions. Observations of business behavior do not support this assertion. On the contrary firms have been observed to raise prices to compensate for increases in overhead costs.^{10/} The answer to the question resides in the corporation's decision processes that determine pricing decisions. If overhead costs are included in this decision process then prices will reflect any increases in fixed costs such as a poll tax.

The point to note in this respect is that it is possible to investigate and determine a firm's pricing policy. In other words, it is no longer either necessary or prudent to rely upon traditional conceptions of firm behavior with the attendant policy prescriptions. For if the classical conclusion on

^{10/} See, for example: W.J. Baumol, Business Behavior, Value and Growth, Macmillan, New York, 1959, p. 78.

poll taxes is in error, and if its conclusions about the means to achieve competitive pricing are open to serious question, the only way in which these issues can be resolved is by empirical investigations of the decisions processes involved. To examine and develop testable theories about economic behavior is not by itself a method for resolving policy disputes. But, it is the only method by which differences of opinion can be compared upon the common ground of empirically valid results. Consequently, if decision processes are to be understood, and if this knowledge is to be used to assist in public and private decision procedures, a considerable effort must be devoted to empirical research. Such research should focus on the economic decision-making processes of individuals and organizations. In addition, these investigations can be carried out in a reasonably systematic way, due to the empirical testability of behavioral hypotheses.

At present, the only impediment to the development of a body of tested hypotheses is a lack of reliable data on individual and organizational decision behavior. This scarcity is not a result of a paucity of ways for collecting it. On the contrary, intensive interviews, detailed observations, and the techniques previously mentioned in the analysis of decision-making behavior are all useful methods for gathering and sorting data. Therefore, since the number of untested hypotheses far exceeds the available data it is toward this endeavor that the emphasis on economic research should be placed.

Chapter 17

TOWARDS A SCIENCE OF ECONOMICS

The primary objective of science is to explain the occurrence of divers phenomena. To carry out this endeavor theories are required which can survive repeated empirical tests. Once developed, such theories perform two vital functions: they provide the theoretical link in the explanatory process; and they are the basis from which empirical knowledge of the phenomena is derived. Without testable theories one can neither explain the occurrence of events nor acquire knowledge about their behavior. In brief, empirical theories are a prerequisite of science.

To be a part of empirical science economics, like any other discipline, must contain theories which can be subjected to the process of disconfirmation by empirical test. Economics as a subject of social study has had a long and distinguished history. As a science, however, it has barely reached the stage of a neophyte. For whether one examines its theoretical statements in the classical or econometric forms, testable hypotheses of economic behavior are no where to be found. Neither of these bodies of theory can be used to explain the occurrence of economic events nor can they be used, in their current formulations, as a basis for the acquisition of empirical knowledge. That this assertion is correct can be seen from the analysis presented in the second and third parts of this book.

Classical mathematical theory fails to produce testable hypotheses because the deductive system ensures that each and every hypothesis contains at least one non-observable equilibrium condition as part of its initial conditions. To submit a hypothesis to an empirically meaningful test, all the relevant initial conditions must be observed, at the onset of the test,

to be empirically true. The presence of equilibria in the initial conditions violates this requirement. For classical theory provides no interpretive rules by which the presence or absence of equilibria can be empirically established. To base a deductive system and, as a consequence, the development of a body of economic theory on the concept of equilibrium can be scientifically successful only if the occurrence of equilibria is readily measurable. However, not only are economic equilibria unobservable, but there is also every reason to believe that no two such points are the same. To construct hypotheses which deal with unobservables makes the task of empirical interpretation difficult enough. To search for general propositions which relate events whose observable characteristics are always changing is to confront the development of a science with insuperable obstacles. In brief, the theories and hypotheses of classical mathematical economics cannot be disconfirmed by empirical test. Consequently, all propositions which are based upon this system, while perhaps being engaging vehicles for after the fact rationalizations, are not a part of empirical science.

In response to this assertion critics may well argue that to be a science is not the whole purpose of economics. Further that classical theory provides an excellent framework with which to analyse and interpret the many important and pressing problems of national and daily economic life. In these situations one frequently wants to know how to act, what to do, or what policy to prescribe. All of these decisions can be taken, so the critics might assert, without any reference to a scientific theory. The principal task is a normative one. If policy prescriptions are carried out in detail, these prescriptions will describe behavior and there is no need to be concerned with explaining previous behavior.

Such a position, however, places the economist in a most unenviable situation. For as long as economic theories are not responsive to data how is one to tell which of a number of competing propositions or prescriptions to follow? Moreover, without the benefit of empirical criticism the acceptance and rejection of theories is based upon their conformity with established positions. In addition, if hypotheses cannot be submitted to test, then it is not possible to develop policy recommendations in the standard engineering way--i.e., to discover by diligent experimental investigation tested methods of achieving the desired results.

Similar difficulties are encountered with econometric theories. In this case it is not the presence of unobservable initial conditions which precludes submitting these hypotheses to empirical test. On the contrary, econometric hypotheses are usually stated so that all variables represent entities which can be directly observed and measured. The empirical obstacles are a consequence of the basic requirements for all statistical tests--i.e., to perform empirically significant statistical tests it must be possible to repeatedly draw samples from the same population.

Econometric hypotheses are stochastic relations. As such their components represent random variables whose populations have certain density functions.

If the "true" characteristics of these populations were known, then statistical tests could be conducted to determine whether particular sample values came from these density functions or not. In actual fact, however, the "true" values are unknown, and sample values are used to calculate parameter estimates. While parameters can be estimated from sample values, the empirical truth or falsity of an estimated relation

can only be determined if it is possible to draw repeated samples from the same population. If this condition cannot be ensured, then such tests as are performed have no empirical significance.

In econometrics the estimated relations are usually tested by the generation of forecasts for the next period. This test on the theory's predictive power is empirically meaningful if and only if the population from which the samples are drawn has remained constant over time. Unfortunately, econometric theory is unable to guarantee that this condition is satisfied. For the only evidence it can produce is whether the forecast is confirmed or not. Clearly, correct or erroneous forecasts can occur for a number of trivial as well as important reasons. The problem is that the occurrence of either does not provide any evidence on the required constancy of the underlying population. Unless and until the stability of such populations can be empirically established it is not possible to subject econometric hypotheses to a process of disconfirmation by empirical test.

This book is concerned with the development of an empirical science of economics. Since neither classical nor econometric theory can satisfy the basic requirements of science, it is necessary to examine recent behavioral formulations to discover whether these theories can provide the requisite empirical foundations.

Behavioral theories represent the decision behavior of a variety of economic actors. A basic premise of these theories is that behavior is a consequence of specific decision processes acting upon the information

available at the time. The information itself may reside in the actor's memory or it may be a product of the environment. In either event such information is describable and refers to observable entities. Hence, it satisfies the condition that the initial conditions be empirically determinable.

The empirical validity of the hypothesized decision processes is less easy to establish. None the less, the procedures by which decision processes are inferred and tested are such as to permit these theories to be confuted by empirical test. As evidence for this statement there are a number of models of economic decision-making behavior which have survived a variety of empirical tests. These tests corroborate both the models themselves as well as the theory of human decision behavior from which they are derived. As a result, behavioral theories can be constructed which satisfy all of the requirements of empirical science.

For behavioral theories to serve as the empirical foundations of economic analysis decision models must be constructed to account for the variety of observable economic behavior. Given such a statement it would appear that a model is required for each behaving unit in the economy. To develop same would clearly be an impractical as well as an exhausting undertaking. Fortunately, it is also unnecessary as the theory of human decision behavior permits the identification of classes of decision behavior which can be encompassed by separate behavioral models.

However, to be able to explain the behavior of a specific class of consumers or firms is not sufficient to satisfy all economists. For there are many who are interested in resolving problems with national or international domains or reference. To develop a science of economics

which can be used by all economists entails the construction of a body of propositions which refer to large aggregates of individual units as well as to the units themselves. Traditionally, econometrics is used when dealing with highly aggregated models while classical economic models are employed when one is working with the individual units. Since neither of these systems can provide testable theories a new approach is required--one which resolves the empirical obstacles while concurrently presenting economists with the desired theoretical tools.

The solution proposed in this book is as follows: Behavioral theories should replace classical theories in all situations where microeconomic analysis is to be employed. With behavioral theories observed decision behavior can be described in detail and explained. The theory's testability permits the economist to acquire empirical knowledge of the behaving units in addition to endowing him with the capability of exploring the consequences of specific policies by experimental investigation.

One further consequence of behavioral theories is that they provide a direct measure of the characteristics of a population's decision processes. That is to say, the minute a model can no longer explain observed behavior techniques are available with which the alterations in the decision processes can be isolated and identified. Such shifts in behavior can be assessed for their effect on the original theory, and by the application of a new set of tests the amended theory can be corroborated or disconfirmed. Clearly, this procedure enables one to detect change and hence measure the stability of specific decision processes. Consequently, behavioral

theories provide a measure on the constancy of decision processes over time with respect to single economic units or collections of such units.

This is the measure that econometric theory lacks. In addition, most econometric theories contain variables which are derived from classical economic theory. Hence, econometric theories, in their current formulations, do not refer to the same processes and variables as are being measured by behavioral theories. If the latter are to serve as the requisite measure of a population's stability over time, the econometric relations must refer to the same processes. The answer is to construct econometric hypotheses in terms of the variables that appear in the discrimination nets of behavioral theories. As long as the process models continue to survive empirical tests, the concomittant econometric relations can be subjected to empirical test. For under these conditions, the stability of the population is assured and the standard statistical test have empirical significance.

The proposal of a solution does not end the matter. For if the major interests of the economist are to be served a large amount of empirical work must be done. In particular, the foundations of behavioral theories must be investigated in greater detail, more models must be constructed and tested in order to develop general theories of specific economic decision processes, and the first of a series of behavioral-econometric theories must be developed and submitted to empirical test. Despite the magnitude of the task, the actual emergence of a science of economics should be by itself a sufficient reward.

85

2-11-67
Date Due

2-11-67		
---------	--	--

Lib-26-67

MIT LIBRARIES



3 9080 003 900 021

83-64

